

(W.I.)

Distribution rights by Oxford Molecular Ltd

t generated

1 MV1SKSP1VL1FSLLSFA.....FLNHQNI NV1NF1VKASH1 42-

Gap 11

225878 seqs, 69334122 residues

Listing first 45 summaries

splitremb1.1.2

13:sp_vertebrate 14:sp_virus

Mean 49.125; Variance 94.306; Scale 0.522

ved by analysis of the total score distribution.

SUMMARIES

2.5	240	10	081256	GLOBULIN-1 (FRAGMENT).	5.92e-69
-----	-----	----	--------	------------------------	----------

ASAY 325
#D1010320

ALIGNMENTS

21	477	12.5	242	10	GBL1271	5.92e-68
22	474	12.4	238	10	GBL0BLIN-1 (FRAGMENT)	2.42e-68
23	470	12.3	242	10	GBL0BLIN-1 (FRAGMENT)	1.58e-67
24	470	12.3	461	10	GBL0BLIN-LIKE PROTEIN	1.58e-67
25	466	12.2	236	10	GBL0BLIN-1 (FRAGMENT)	1.03e-66
26	463	12.2	238	10	GBL1250	4.20e-66
27	464	12.2	242	10	GBL0BLIN-1 (FRAGMENT)	2.63e-66
28	466	12.2	407	10	GBL1250	1.03e-66
29	448	11.8	423	10 ⁶	GBL1750	4.64e-65
30	443	11.4	421	10	PHASEOLIN PRECURSOR	4.99e-60
31	429	11.3	428	10	PHASEOLIN PRECURSOR	3.19e-55
32	426	11.2	414	10	PHASEOLIN, 7S GLOBULIN	1.28e-55
33	418	11.0	430	10	PHASEOLIN PRECURSOR	5.21e-55
34	418	11.0	430	10	ALPHA- <i>PHASEOLIN</i> PRECUR	5.21e-55
35	391	10.3	540	10	GBL0BLIN-1 S ALLELE PR	1.30e-51
36	324	8.5	504	10	GBL0371	1.90e-38
37	318	8.4	186	10	GBL0371	2.73e-37
38	319	8.4	287	10	GBL0376	1.75e-33
39	286	7.5	240	10	7S STORAGE PROTEIN ALP	3.56e-31
40	265	7.0	136	10	GBL0BLIN (FRAGMENT)	3.13e-22
41	241	6.3	115	10	VICILIN-LIKE PROTEIN	8.32e-22
42	230	6.0	118	10	BETA-CONGLUYCININ-ALPHA	8.21e-21
43	205	5.4	83	10	GBL0222	2.26e-16
44	204	5.4	159	10	GBL0BLIN-LIKE PROTEIN	3.38e-16
45	199	5.2	171	11	GLUTAMINE REPEAT PROTE	2.50e-15

ALIGNMENTS

RESULT	1	PRELIMINARY:	PRT:	525 AA.
AC	043358			
AC	043358;			
DT	01-NOV-1996 (TREMBLrel. 01, Created)			
DT	01-NOV-1996 (TREMBLrel. 01, Last sequence update)			
DT	01-NOV-1999 (TREMBLrel. 12, Last annotation update)			
DE	VICILIN PRECURSOR.			
GN	CSV.			
OS	Theobroma=cacao*(Cacao).			
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;			
OC	euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;			
OC	core eudicots; Rosidae; eurosids II; Malvales; Malvaceae; Theobroma.			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE-LEAVES:			
RX	MEDLINE; 92288309.			
RA	MCHEHRY, J., FRITZ, J.,			
RT	Comparison of the structure and nucleotide sequences of vicilin genes			
RT	of cocoa and cotton raise questions about vicilin evolution.";			
RL	Plant Mol. Biol. 18,1173-1176(1992).			
DR	EMBL; X62625; CAA44493.1;			
DR	EMBL; X62626; CAA44494.1;			
DR	HSSP; P02853; 2PHL.			
DR	MENDEL; 30919; Thecc;1188;30919.			
DR	PRAM; PF00546; Seedstore_7s; 1.			
KM	Signal.			
FT	CHAIN	1	24	POTENTIAL.
FT	CHAIN	25	525	VICILIN
SO	SEQUENCE	525 AA;	60798 MW;	4890ENCE CRC32;

```
Matches: 525; Conservative 0; Mismatches 0; Indels 0; Gaps 0
```

3030

Db 1 MWTSKSFYLLISLLLSFALLCSGSAVGAKRYENDPQKQIQEGRKSESETEEREDE 60
0Y 1 MWTSKSFYLLISLLLSFALLCSGSAVGAKRYENDPQKQIQEGRKSESETEEREDE 60
Db 61 QGSEGRCEREYKEDQROOEELQKQYQCCGRCCEQDQGGREDDQCCQKCKMEYKEDERGE 120
0Y 61 QCEGRCEREYKEDQROOEELQKQYQCCGRCCEQDQGGREDDQCCQKCKMEYKEDERGE 120
Db 121 HENYHHKKKRRSEEGCGQRRNNPYTPPKRRSFOTPRDEGQFKILQRAENSPLKGIN 180

(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

Distribution rights by Oxford Molecular Ltd

```
Mpsrch_pp protein - protein database search, using Smith-Waterman algorithm
Run on: Sat May 13 11:35:26 2000; MspPar time 66.80 Seconds
101.841 Million cell updates/sec
```

Tabular output not generated

Title: >US-09-331-631-7
Description: (1-525) from US09331631 pep (1 of 3)

Perfect Score: 3808
Sequence: 1 MWISKSPFIVLIFSLLSFA.....FLNHKQNTNVIKFTVKASAY 5229

Scoring table: PAM 150
Gap 11

Searched: 131253 seqs, 12956647 residues

Post-processing: Minimum Match 08
Listing first 45 summaries

```
Database: a-issued
1.5A COMB 3.5B COMB 3.0CT COMB 4.backfiles
```

Statistics: Mean 33.485; Variance 165.024; scale 0.203

Pred. NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysts of the total score distribution.

SUMMARIES

No.	Score	Query	Match	Length	DB	ID	Description	Pred. No.
1	3654	96.0	566	1	US-07-955-	Sequence 2, Applicatio	0.00e+00	
2	3654	96.0	566	1	US-07-955-	Sequence 22, Applicati	0.00e+00	
3	1858	48.0	587	1	US-07-955-	Sequence 21, Applicati	1.70e-11	
4	805	21.1	605	1	US-07-955-	Sequence 24, Applicati	2.12e-55	
5	776	20.4	571	1	US-07-955-	Sequence 25, Applicati	6.16e-55	
6	732	19.2	410	1	US-07-955-	Sequence 26, Applicati	3.32e-55	
7	434	11.4	421	1	US-07-955-	Sequence 27, Applicati	0.00e+00	
8	173	4.5	1162	2	US-08-728-	Sequence 2, Applicatio	1.62e-00	
9	164	4.3	1898	2	US-08-800-	Sequence 94, Applicati	7.46e-04	
10	164	4.3	1898	1	US-08-056-	Sequence 94, Applicati	7.46e-04	
11	147	3.9	788	2	US-08-918-	Sequence 43, Applicati	1.27e-07	
12	139	3.7	186	2	US-08-918-	Sequence 4, Applicatio	4.73e-07	
13	120	3.2	303	1	US-08-456-	Sequence 2, Applicatio	9.90e-07	
14	120	3.2	303	2	US-08-456-	Sequence 2, Applicatio	9.90e-07	
15	120	3.2	303	1	US-08-109-	Sequence 2, Applicatio	9.90e-07	
16	122	3.2	428	1	US-08-193-	Sequence 29, Applicati	7.23e-01	
17	121	3.2	432	2	US-08-935-	Sequence 47, Applicati	8.46e-01	
18	123	3.2	1311	2	US-08-540-	Sequence 4, Applicatio	6.17e-01	
19	123	3.2	1311	3	PCI-US95-1	Sequence 4, Applicatio	6.17e-01	
20	115	3.0	344	4	US210183-1	Patent No. 5210183.	2.16e+00	
21	115	3.0	683	4	5210183-3	Patent No. 5210183.	2.16e+00	
22	111	2.9	376	4	5180810-1	Patent No. 5180810.	4.00e+00	
23	110	2.9	816	2	US-08-267-	Sequence 9, Applicatio	4.66e+00	

LONG

ALIGNMENTS

24	110	2.9	2703	1	US-08-184-	Sequence 19,	Applicati	4.66e+00
25	108	2.8	162	1	PCT-US93-0	Sequence 3,	Applicatio	6.33e+00
26	108	2.8	678	3	PCT-US93-0	Sequence 3,	Applicatio	6.33e+00
27	106	2.8	903	2	US-08-853-	Sequence 5,	Applicatio	8.57e+00
28	104	2.7	303	1	US-08-183-	Sequence 2,	Applicatio	1.16e+01
29	103	2.7	360	2	US-08-531-	Sequence 2,	Applicatio	1.35e+01
30	103	2.7	737	1	US-08-183-	Sequence 2,	Applicatio	1.16e+01
31	104	2.7	737	1	US-08-185-	Sequence 4,	Applicatio	2.11e+01
32	100	2.6	614	3	PCT-US95-0	Sequence 21,	Applicati	5.08e+01
33	94	2.5	78	2	US-08-465-	Sequence 53,	Applicati	5.08e+01
34	94	2.5	162	2	US-08-485-	Sequence 63,	Applicati	5.08e+01
35	94	2.5	162	2	US-08-461-	Sequence 63,	Applicati	5.08e+01
36	97	2.5	361	1	US-08-415-	Sequence 4,	Applicatio	4.38e+01
37	95	2.5	542	3	PCT-US92-1	Sequence 13,	Applicati	4.39e+01
38	95	2.5	542	3	PCT-US92-1	Sequence 13,	Applicati	4.39e+01
39	95	2.5	542	1	US-07-814-	Sequence 8,	Applicatio	4.39e+01
40	95	2.5	542	1	US-08-325-	Sequence 14,	Applicati	4.39e+01
41	95	2.5	559	1	US-08-314-	Sequence 9,	Applicati	4.39e+01
42	96	2.5	905	2	US-08-574-	Sequence 7,	Applicatio	3.80e+01
43	96	2.5	1135	2	US-08-574-	Sequence 28,	Applicati	3.80e+01
44	96	2.5	1876	2	US-08-609-	Sequence 12,	Applicati	3.80e+01
45	96	2.5	1876	2	US-08-609-	Sequence 12,	Applicati	3.80e+01

Sequence 2, Application US/07955905A

Sequence 2, Application US/07955905A

GENERAL INFORMATION:

APPLICANT:

TITL E OF INVENTION

NUMBER OF SEQUENCES
COMPUTER READABLE

MEDIUM TYPE: F1

OPERATING SYSTEM

CURRENT APPLICATION

APPLICATION NUMBER: 21
 FILING DATE: 21

CLASSIFICATION:

SEQUENCE CHARACTER

LENGTH: 200 amino acids
TYPE: amino acid

TOPOLOGY: linear

SEQUENCE 566 AA; 656

96 Every Match

504: Conservat

.....

1 MYJ3K3F1V6LFSUEE

61 QCEQRCEREYKEQQROU

61 QCEQRCEREYKEQQRQ

121 HENYHNHKKNRSEEEEG

UY 14: HENYHKKNSSEEGOONNPPYFPKRSFOTRDEGNFKILORFAENSPPLKGIN 180
Db 181 DYRLAMEANPNFTILPHHCDAEAIYVTNGKGTITFVTHNKESYVNGRTVSVAPS 240
OY 181 DYRLAMEANPNFTILPHHCDAEAIYVTNGKGTITFVTHNKESYVNGRTVSVAPS 240
Db 241 TVYVSDNOEKLITAVLALPVNSPGKYLEFFPAGNNKPESYGAASYLETFTORE 300
OY 241 TVYVSDNOEKLITAVLALPVNSPGKYLEFFPAGNNKPESYGAASYLETFTORE 300
Db 301 KLEELIEEOGRKROOGOCGMFRKAKPEOIRAIISOATSPRRHGERLAINLLSOPVYS 360
OY 301 KLEELIEEOGRKROOGOCGMFRKAKPEOIRAIISOATSPRRHGERLAINLLSOPVYS 360
Db 361 NONGRFEACPEDEFSOFQNDVAVSAFKLNOGAIFFVPHNSKATFVFTDGYAOMAC 420
OY 361 NONGRFEACPEDEFSOFQNDVAVSAFKLNOGAIFFVPHNSKATFVFTDGYAOMAC 420
Db 421 PHLSROSOGSGRODRREDEESEEETFGFEQVAKAPLSPGDVFAVAGHATFFASND 480
OY 421 PHLSROSOGSGRODRREDEESEEETFGFEQVAKAPLSPGDVFAVAGHATFFASND 480
Db 481 OPLNAVAFGLNANNNORIFLAGRPFLNHKN-TNVIKFTV 520
OY 481 OPLNAVAFGLNANNNORIFLAGRPFLNHKN-TNVIKFTV 520
RESULT 2
ID US-07-955-905A-22 STANDARD: PRT: 566 AA.
XX xxxxxx
AC xxxxxx
DE Sequence 22, Application US/07955905A
XX Sequence 22, Application US/07955905A
CC Patent No. 5770433
CC GENERAL INFORMATION:
CC APPLICANT:
CC TITLE OF INVENTION: RECOMBINANT 47 AND 31 KD COCOA PROTEINS AND
CC NUMBER OF SEQUENCES: 28
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC OPERATING SYSTEM: IBM PC compatible
CC SOFTWARE: Patentin Release #1.0, Version #1.25 (EPO)
CC CURRENT APPLICATION DATA:
CC FILING DATE: 21-JAN-1993
CC CLASSIFICATION: 435
CC INFORMATION FOR SEQ ID NO: 22:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 566 amino acids
CC TOPOLOGY: linear
CC MOLECULE TYPE: protein
CC ORGANISM: Theobroma cacao
CC FEATURE:
CC NAME/KEY: Protein
CC LOCATION: 1-566
CC OTHER INFORMATION: /note- "97 kd precursor protein"
SO SEQUENCE 566 AA: 65613 MW: 1529150 CN:
Query Match 96.0% Score 3654: DB 1: Length 566:
Best Local Similarity 96.7% Pred. No. 0.00e+00:
Matches 504: Conservative 10: Mismatches 6: Indels 1: Gaps 1:
1 MYSKSPFIVLIFSLFALLCGSAVGRKQYERPROOYECQRCSEATEEREDE 60
NYSKSPFIVLIFSLFALLCGSAVGRKQYERPROOYECQRCSEATEEREDE 60

Ub 61 QCEORCERETKEEOQKQDEELQYQOCQGRCOEQQOQRCCKMEQYKEDEGE 120
OY 61 QCEORCERETKEEOQKQDEELQYQOCQGRCOEQQOQRCCKMEQYKEDEGE 120
Db 121 HENYHKKNSSEEGOONNPPYFPKRSFOTRDEGNFKILORFAENSPPLKGIN 180
OY 121 HENYHKKNSSEEGOONNPPYFPKRSFOTRDEGNFKILORFAENSPPLKGIN 180
Db 181 DYRLAMEANPNFTILPHHCDAEAIYVTNGKGTITFVTHNKESYVNGRTVSVAPS 240
OY 181 DYRLAMEANPNFTILPHHCDAEAIYVTNGKGTITFVTHNKESYVNGRTVSVAPS 240
Db 241 TVYVSDNOEKLITAVLALPVNSPGKYLEFFPAGNNKPESYGAASYLETFTORE 300
OY 241 TVYVSDNOEKLITAVLALPVNSPGKYLEFFPAGNNKPESYGAASYLETFTORE 300
Db 301 KLEELIEEOGRKROOGOCGMFRKAKPEOIRAIISOATSPRRHGERLAINLLSOPVYS 360
OY 301 KLEELIEEOGRKROOGOCGMFRKAKPEOIRAIISOATSPRRHGERLAINLLSOPVYS 360
Db 361 NONGRFEACPEDEFSOFQNDVAVSAFKLNOGAIFFVPHNSKATFVFTDGYAOMAC 420
OY 361 NONGRFEACPEDEFSOFQNDVAVSAFKLNOGAIFFVPHNSKATFVFTDGYAOMAC 420
Db 421 PHLSROSOGSGRODRREDEESEEETFGFEQVAKAPLSPGDVFAVAGHATFFASND 480
OY 421 PHLSROSOGSGRODRREDEESEEETFGFEQVAKAPLSPGDVFAVAGHATFFASND 480
Db 481 OPLNAVAFGLNANNNORIFLAGRPFLNHKN-TNVIKFTV 520
OY 481 OPLNAVAFGLNANNNORIFLAGRPFLNHKN-TNVIKFTV 520
RESULT 3
ID US-07-955-905A-23 STANDARD: PRT: 587 AA.
XX xxxxxx
AC xxxxxx
DE Sequence 23, Application US/07955905A
XX Sequence 23, Application US/07955905A
CC Patent No. 5770433
CC GENERAL INFORMATION:
CC APPLICANT:
CC TITLE OF INVENTION: RECOMBINANT 47 AND 31 KD COCOA PROTEINS AND
CC NUMBER OF SEQUENCES: 28
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC OPERATING SYSTEM: IBM PC compatible
CC SOFTWARE: Patentin Release #1.0, Version #1.25 (EPO)
CC CURRENT APPLICATION DATA:
CC FILING DATE: 21-JAN-1993
CC CLASSIFICATION: 435
CC INFORMATION FOR SEQ ID NO: 23:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 587 amino acids
CC TOPOLOGY: linear
CC MOLECULE TYPE: protein
CC ORGANISM: Gossypium hirsutum
CC FEATURE:
CC NAME/KEY: Protein
CC LOCATION: 1-587
CC OTHER INFORMATION: /note- "Vicilin from G. hirsutum"
SO SEQUENCE 587 AA: 69497 MW: 1625753 CN:

 W O R L D
 (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit,
 Copyright (c) 1993-1998 University of Edinburgh, U.K.
 Distribution rights by Oxford Molecular Ltd

MParch_pp protein - protein database: MARCH, using Smith-Waterman algorithm
 Run on: Sat May 13 11:45:14 2000; Maspar time 58.52 Seconds
 130.618 Million cell updates/sec

Tabular output not generated.

Title: >US-09-331-631-8
 Description: (1-599) from US09331631: pep (1 of 4)
 Perfect Score: 4332
 Sequence: 1 MVRKRSACVLLFSLFSLFG.....FNSNPQESFVSRORQASE 590

Scoring table: PAM 150
 Gap 11

Searched: 131253 seqs, 12956647 residues

Post-processing: Minimum Match 0%
 Listing first 45 summaries

Database: a-issued
 1:5A_COMB 2:5B_COMB 3:PCT_COMB 4:backfillset

Statistics: Mean 34.226; Variance 174.925; scale 0.196

Pred. No. is the number of results predicted by chance to have a
 score greater than or equal to the score of the result being printed,
 and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description	Pred. No.
1	4119	95.1	587	1	US-07-955- Sequence 23, Applicati	0.00e+00
2	2101	48.5	566	1	US-07-955- Sequence 22, Applicati	7.38e-162
3	2101	48.5	566	1	US-07-955- Sequence 2, Applicati	7.38e-162
4	968	22.3	605	1	US-07-955- Sequence 24, Applicati	1.71e-67
5	832	19.2	571	1	US-07-955- Sequence 25, Applicati	2.62e-56
6	776	17.9	410	1	US-07-955- Sequence 26, Applicati	1.01e-51
7	493	11.4	421	1	US-07-955- Sequence 27, Applicati	8.72e-29
8	223	5.1	1162	2	US-08-728- Sequence 2, Applicati	8.99e-08
9	166	3.8	1898	2	US-08-800- Sequence 94, Applicati	1.23e-03
10	166	3.8	1898	2	US-08-056- Sequence 94, Applicati	1.23e-03
11	162	3.7	788	2	US-08-918- Sequence 4, Applicati	2.33e-03
12	138	3.2	186	2	US-08-557- Sequence 43, Applicati	1.07e-01
13	127	2.9	361	1	US-08-415- Sequence 4, Applicati	5.84e-01
14	123	2.8	737	1	US-08-185- Sequence 4, Applicati	1.07e-00
15	123	2.8	737	1	US-08-185- Sequence 2, Applicati	1.07e-00
16	119	2.7	434	1	US-08-337- Sequence 3, Applicati	1.96e+00
17	119	2.7	2237	2	US-08-223- Sequence 48, Applicati	1.96e+00
18	119	2.7	2237	2	US-08-455- Sequence 47, Applicati	1.96e+00
19	119	2.7	2339	2	US-08-223- Sequence 47, Applicati	1.96e+00
20	119	2.7	2339	2	US-08-910- Sequence 2, Applicati	5.57e+00
21	112	2.6	357	1	US-08-552- Sequence 2, Applicati	5.57e+00
22	112	2.6	357	1	US-08-552- Sequence 2, Applicati	5.57e+00
23	112	2.6	357	3	PCT-US95-0 Sequence 2, Applicati	5.57e+00

24	112	2.6	369	2	US-08-474- Sequence 32, Applicati	5.57e+00
25	112	2.6	369	1	US-07-688- Sequence 32, Applicati	5.57e+00
26	112	2.6	369	3	PCT-US91-0 Sequence 31, Applicati	5.57e+00
27	112	2.6	914	1	US-08-484- Sequence 2, Applicati	5.57e+00
28	112	2.6	914	1	US-08-484- Sequence 42, Applicati	5.57e+00
29	112	2.6	3144	2	US-08-457- Sequence 6, Applicati	5.57e+00
30	112	2.6	3144	1	US-08-246- Sequence 6, Applicati	5.57e+00
31	112	2.6	3144	1	US-08-453- Sequence 9, Applicati	8.65e+00
32	109	2.5	156	3	PCT-US95-0 Sequence 9, Applicati	8.65e+00
33	109	2.5	156	1	US-08-552- Sequence 9, Applicati	8.65e+00
34	109	2.5	156	1	US-08-552- Sequence 9, Applicati	8.65e+00
35	110	2.5	240	3	PCT-US95-0 Sequence 73, Applicati	7.47e+00
36	110	2.5	240	3	PCT-US95-0 Sequence 2, Applicati	8.65e+00
37	109	2.5	303	2	US-08-459- Sequence 2, Applicati	8.65e+00
38	109	2.5	303	2	US-08-460- Sequence 2, Applicati	8.65e+00
39	109	2.5	303	1	US-08-109- Sequence 5, Applicati	1.16e+01
40	107	2.5	303	1	US-08-185- Sequence 15, Applicati	8.65e+00
41	109	2.5	356	1	US-08-552- Sequence 15, Applicati	8.65e+00
42	109	2.5	356	1	US-08-910- Sequence 15, Applicati	8.65e+00
43	109	2.5	387	2	US-08-991- Sequence 2, Applicati	1.16e+01
44	107	2.5	387	1	US-08-539- Sequence 2, Applicati	1.16e+01
45	107	2.5	387	1	US-08-529- Sequence 2, Applicati	1.16e+01

ALIGNMENTS

RESULT 1
 ID US-07-955-905A-23 STANDARD: PRT: 587 AA.
 XX xxxxxx
 XX
 AC
 AC
 DT
 XX
 XX
 DE Sequence 23, Application US/07955905A
 CC Sequence 23, Application US/07955905A
 CC Patent No. 5770433
 CC GENERAL INFORMATION:
 CC APPLICANT:
 CC TITLE OF INVENTION: RECOMBINANT 47 AND 31 KD COCOA PROTEINS AND
 CC TITLE OF INVENTION: PRECURSOR
 CC NUMBER OF SEQUENCES: 28
 CC COMPUTER READABLE FORM:
 CC MEDIUM TYPE: Floppy disk
 CC COMPUTER: IBM PC compatible
 CC OPERATING SYSTEM: PC-DOS/MS-DOS
 CC SOFTWARE: Patentin Release #1.0, Version #1.25 (EPO)
 CC CURRENT APPLICATION DATA:
 CC APPLICATION NUMBER: US/07/955,905A
 CC FILING DATE: 21-JAN-1993
 CC CLASSIFICATION: 435
 CC INFORMATION FOR SEQ ID NO: 23:
 CC SEQUENCE CHARACTERISTICS:
 CC LENGTH: 587 amino acids
 CC TYPE: amino acid
 CC TOPOLOGY: linear
 CC MOLECULE TYPE: protein
 CC ORGANISM: Gossypium hirsutum
 CC FEATURE:
 CC NAME/KEY: Protein
 CC LOCATION: 1..587
 CC OTHER INFORMATION: /note="Vicilin from G. hirsutum"
 CC SEQUENCE 587 AA: 69497 MW: 1625753 CN:

Query Match 95.1%; Score 4119; DB 1; Length 587;
 Best Local Similarity 95.1%; Pred. No. 0.00e+00;
 Matches 561; Conservative 17; Mismatches 9; Indels 3; Gaps 2;
 Db 1 MVRKRSACVLLFSLFSLGCSANDPGRKDDDPKRYEDCRRRGEWDTRGQKQOQ 60
 1 MVRKRSACVLLFSLFSLGCSAKDFPGRGDDDPKRYEDCRRRCEWDTGQKQOQ 60

NO.	Score	Match	Length	DB	ID	Description	freq. n
1	3654	96.0	566	2	S32477	vicillin precursor - c	0.006+00
2	1919	50.4	505	2	S06398	alpha-globulin type A	0.006+00
3	1910	50.2	509	2	S08059	alpha-globulin type B	0.006+00
4	1853	48.7	588	1	FWCMB	alpha-globulin B prec	0.006+00
5	881	21.1	573	3	A53234	globulin-15, GB15 -	3.82e-11
6	830	22.3	582	2	B53234	vicillin-like storage	1.82e-11
7	848	21.8	582	2	S00567	vicillin precursor (cl	7.85e-11
8	827	21.7	439	1	FWSYCB	beta-conglycinin beta	3.17e-11
9	826	21.7	459	1	S08505	vicillin - garden pea	5.04e-11
10	810	21.3	605	1	FWSYBA	beta-conglycinin alph	8.49e-11
11	805	21.1	605	2	S20007	beta-conglycinin alph	8.64e-11
12	776	20.4	571	2	S00566	convicillin precursor	5.92e-11
13	771	20.2	463	2	A77288	vicillin precursor - f	5.99e-11
14	771	20.2	463	2	S06309	vicillin precursor (cl	5.99e-11
15	764	20.1	438	2	S37577	vicillin, 47K - garden	1.53e-11
16	732	19.2	410	1	FMPMB	vicillin B precursor -	3.98e-11
17	724	19.0	448	1	S47456	vicillin-like storage	1.59e-11
18	702	18.4	524	2	J01730	6K sucrose-binding p	3.95e-11
19	697	18.3	483	2	T06459	6K sucrose-binding p	3.94e-11
20	682	17.5	450	2	S15675	globulin-2 precursor	3.84e-11
21	644	16.9	445	2	S00281	canaavallin - sword bea	1.37e-99
22	643	16.9	445	2	J02264	canaavallin - jack bean	2.16e-99
23	634	16.6	639	2	B34810	beta-conglycinin alph	1.33e-99

24	593	15.6	637	2	S55221	globoin	Begl	precursor	1.65e-87
25	531	13.9	196	2	S02281	convicillin	(clone pcd		2.33e-75
26	531	13.9	196	2	S02281	convicillin	(clone pcd		2.33e-75
27	470	12.5	236	2	T01662	globoblin-1	- maize (f		1.07e-64
28	466	12.2	436	2	A23498	phasellon	type alpha		1.53e-63
29	435	11.4	275	1	FPMWA	globoblin	- maize		8.98e-63
30	435	11.4	275	1	FPMWA	vicillin A precursor			7.75e-57
31	446	11.2	421	2	A24810	phasellon beta chain			1.20e-56
32	418	11.0	430	2	S60190	vicillin - zambia	farfu		4.01e-55
33	391	10.3	540	2	S21825	alpha-phasellon	precu		1.33e-53
34	365	9.6	296	2	T06372	vicillin-like storage			1.67e-48
35	286	7.5	240	2	T06399	convicillin precursor			1.23e-43
36	230	6.0	218	1	S16334	7S storage protein	precursor		3.33e-28
37	222	5.8	124	1	FPMW4	beta-conglycinin	alpha		1.98e-15
38	195	5.1	613	2	S27770	vicillin, 14k componen			4.48e-18
39	181	5.0	136	2	T02870	hypochloral protein			1.31e-13
40	168	4.9	184	1	FSEB	globulin 2 precursor			5.81e-11
41	183	4.8	1038	2	T02634	phasellon - kidney be			1.76e-12
42	170	4.5	395	2	S35565	rep protein homolog			1.11e-11
43	170	4.5	395	2	S35565	sex-determining prote			1.21e-07
44	173	4.5	708	2	A53185	sex-determining prote			1.21e-07
45	171	4.5	1094	2	S59313	G-box-binding factor			4.13e-10
						protein kinase - blm			8.45e-10

ALIGNMENTS

ENTRY	1
TITLE	
ORGANISM	
DATE	
ACCESSIONS	
REFERENCE	
#authors	
#journal	
#title	
S22477	#type complete
vicilin precursor - cacao	
#formal_name Theobroma cacao	#common_name cacao
30-Sep-1993	sequence_revision 30-Sep-1993
17-Mar-1999	#text_change
S22477; S22478; S18105; S22050	
S22477	
McHenry, L.; Fritze, P.J.	
Plant Mol. Biol. (1992) 18:1173-1176	
Comparison of the structure and nucleotide sequence of	
vicilin genes of cocoa and cotton raise questions about	
vicilin evolution.	

```

#accession      S22477
#molecule_type DNA
#residues       1-566  ##label MCH
#cross-references EMBL:X62625
#accession      S22478

```

GENETICS 2007 2013 2002 2003

CLASSIFICATION *superfamily glycinin

KEYWORDS seed; storage protein

```

1-24      #domain signal sequence #status predicted #label SIC
25-566    #product; vicillin #status predicted #label MAT
SUMMARY    #length 566 #molecular-weight 65513 #checksum 3693

```

Query Match	96.0%;	Score 3654;	DB 2;	length 566;
Best Local Similarity	96.7%;	Pred. NO. 0.00e+00;		
Matches	504;	Conservative	10;	Mismatches 6; Indels 1; Gaps 1

Db	1	MYISKSPFYILITFLLSFALLCSGVSAVGRKQYERPRDQYEECCQRCESEALFEENDE	60
	1	MYISKSPFYILITFLLSFALLCSGVSAVGRKQYERPRDQYEECCQRCESEALFEENDE	60
Qy	61	OCQRCREKVEKQOROOEEELORQYOCQRCQEOQOREOQOCQKCMQYKEORGE	120
Db	61	OCQRCREKVEKQOROOEEELORQYOCQRCQEOQOREOQOCQKCMQYKEORGE	120
Qy	61	OCQRCREKVEKQOROOEEELORQYOCQRCQEOQOREOQOCQKCMQYKEORGE	120
Db	121	HEVYHNHKKRRSEEEQOQRNNPYFPFKRRSFQTRFDEEGNFKILLORFANSPPLGIN	180
Qy	121	HEVYHNHKKRRSEEEQOQRNNPYFPFKRRSFQTRFDEEGNFKILLORFANSPPLGIN	180
Qy	121	HEVYHNHKKRRSEEEQOQRNNPYFPFKRRSFQTRFDEEGNFKILLORFANSPPLGIN	180

Db 181 DYRLAMFANPTFLPHHCOAEIYVINGKTIFFVTHEKESYNORGTWVSPAGS 240
181 DYRLAMFANPTFLPHHCOAEIYVINGKTIFFVTHEKESYNORGTWVSPAGS 240
Db 241 TVYVVSQDNOEKLTIYALVLPVNSPGKYLEFFPAGNPKPESTYGAFTVETVNTORE 300
241 TVYVVSQDNOEKLTIYALVLPVNSPGKYLEFFPAGNPKPESTYGAFTVETVNTORE 300
Db 301 KLEELLEDROGKROGOOGGFRKAKPEOIRAIISOATSPRRHGERLAINLLSOPYS 360
301 KLEELLEDROGKROGOOGGFRKAKPEOIRAIISOATSPRRHGERLAINLLSOPYS 360
Db 361 NONGRFFECAPEDFQFONMDVAVSAFKLNOGAIFVPHYSKATFVVTDOGYAOMAC 420
361 NONGRFFECAPEDFQFONMDVAVSAFKLNOGAIFVPHYSKATFVVTDOGYAOMAC 420
Db 421 PHLSQOSGOSGRODRREOESEETFGFQVAKPLSPDVAVAPAGHATVFFASKD 480
421 PHLSQOSGOSGRODRREOESEETFGFQVAKPLSPDVAVAPAGHATVFFASKD 480
Db 481 OPLNVAAGLNAONNORIFLAGKMLVROMDSEAKELSGV 521
481 OPLNVAAGLNAONNORIFLAGKMLVROMDSEAKELSGV 521
OY 481 OPLNVAAGLNAONNORIFLAGKMLVROMDSEAKELSGV 520
RESULT 2
ENTRY 506398 #type complete
TITLE alpha-globulin type A precursor - upland cotton
ALTERNATE_NAMES seed storage protein
ORGANISM #formal_name Gossypium hirsutum #common_name upland cotton
DATE 31-Mar-1990 #sequence_revision 31-Mar-1990 #text_change
ACCESSIONS 506398
REFERENCE 506398
#authors Chien, C.A.; Borroto, K.; Kamalay, J.A.; Dure III, L.
#journal Plant Mol. Biol. (1987) 9:533-546
#title Developmental biochemistry of cottonseed embryogenesis and germination. XIX. Sequences and genomic organization of the alpha globulin (vicilin) genes of cottonseed.
#accession 506398
#status not compared with conceptual translation
#molecule_type DNA
#residues 1-605 #label CHL
CLASSIFICATION #superfamily glycylin
FEATURE 1-24
25-605 #domain signal sequence #status predicted #label SIG
#product alpha-globulin type A #status predicted #label MAT
SUMMARY #length 605 #molecular_weight 71049 #checksum 4376
Query Match 50.48: Score 1919; DB 2; Length 605;
Best Local Similarity 57.08; Pred. No. 0.00e+00;
Matches 269; Conservative 101; Mismatches 87; Indels 15; Gaps 13;
Db 79 DPOARRVODCRHQOE--ERLRPHCEQSCREY--EKQOQOQPD--KQFKECOQRCOWOE 133
79 DPOARRVODCRHQOE--ERLRPHCEQSCREY--EKQOQOQPD--KQFKECOQRCOWOE 133
OY 37 DPOQVQEQOCORCESEATEEREEOQEOECERCEYKQEQOEEELORQVQOCGRQCO--EQ 95
37 DPOQVQEQOCORCESEATEEREEOQEOECERCEYKQEQOEEELORQVQOCGRQCO--EQ 95
Db 134 OPPER--KOQCVRECRYOYEDPMKGERENKMPREEEEDDEQOQRNNPYFH--RRSQ 191
134 OPPER--KOQCVRECRYOYEDPMKGERENKMPREEEEDDEQOQRNNPYFH--RRSQ 191
OY 96 COGQRCQOCCKCKCHQYKEOE--RGEHEN--YHNHKNRSEEEGQQRNNPYFPKRSPQ 153
96 COGQRCQOCCKCKCHQYKEOE--RGEHEN--YHNHKNRSEEEGQQRNNPYFPKRSPQ 153
Db 192 EPRFEHGNFVLOFPAKHLRGINFERIALILEANPTFVLPHHCDAEITYVYVINGSG 251
192 EPRFEHGNFVLOFPAKHLRGINFERIALILEANPTFVLPHHCDAEITYVYVINGSG 251
OY 154 TFRDEGCFKILQFPAKSPPLKGINDRFLAMFENPTFLPHHCDAEITYVYVINGSG 213
154 TFRDEGCFKILQFPAKSPPLKGINDRFLAMFENPTFLPHHCDAEITYVYVINGSG 213
Db 252 TATFTTHEKESYNVPGVYVIRIPAGSTYVLANODNREKLTIAVLRPVNPGOFQKFFP 311
252 TATFTTHEKESYNVPGVYVIRIPAGSTYVLANODNREKLTIAVLRPVNPGOFQKFFP 311
OY 216 TITFTTHEKESYNVPGVYVIRIPAGSTYVLANODNREKLTIAVLRPVNPGOFQKFFP 273
216 TITFTTHEKESYNVPGVYVIRIPAGSTYVLANODNREKLTIAVLRPVNPGOFQKFFP 273
Db 370 RQENRQSLRILFRELLEAVNTRESEOLDELPGRQSRHROGO--GMFRKASQOIRAL 370
370 RQENRQSLRILFRELLEAVNTRESEOLDELPGRQSRHROGO--GMFRKASQOIRAL 370
OY 333 RQENRQSLRILFRELLEAVNTRESEOLDELPGRQSRHROGO--GMFRKASQOIRAL 333
333 RQENRQSLRILFRELLEAVNTRESEOLDELPGRQSRHROGO--GMFRKASQOIRAL 333

Db 371 SQGATSPKGSSEGTAFNLLSOTPRYSNOKRFEACPRNFQOOLREVDSVAFELINKG 430
371 SQGATSPKGSSEGTAFNLLSOTPRYSNOKRFEACPRNFQOOLREVDSVAFELINKG 430
OY 334 SQGATSPKGSSEGTAFNLLSOTPRYSNOKRFEACPRNFQOOLREVDSVAFELINKG 392
334 SQGATSPKGSSEGTAFNLLSOTPRYSNOKRFEACPRNFQOOLREVDSVAFELINKG 392
Db 431 SIFPHYSKATFVYVTEGNGEHCPCPHLSROSSD--SREEE--EOEEVEVRSQY 488
431 SIFPHYSKATFVYVTEGNGEHCPCPHLSROSSD--SREEE--EOEEVEVRSQY 488
OY 393 AIFPHYSKATFVYVTEGNGEHCPCPHLSROSSD--SREEE--EOEEVEVRSQY 452
393 AIFPHYSKATFVYVTEGNGEHCPCPHLSROSSD--SREEE--EOEEVEVRSQY 452
Db 489 KRYRAQSTGNLFFVYPAGHPYTFVASONEDLLGLGFLYNGQDNKRIFVACK 540
489 KRYRAQSTGNLFFVYPAGHPYTFVASONEDLLGLGFLYNGQDNKRIFVACK 540
OY 453 QVYAPLSPGDVAVAPAGHATVFFASKDQPLNVAAGL--MNONNORITLAOR 503
453 QVYAPLSPGDVAVAPAGHATVFFASKDQPLNVAAGL--MNONNORITLAOR 503
RESULT 3
ENTRY 508059 #type fragment
TITLE alpha-globulin type B precursor (tandem 1) - upland cotton
ALTERNATE_NAMES seed storage protein
ORGANISM #formal_name Gossypium hirsutum #common_name upland cotton
DATE 31-Mar-1990 #sequence_revision 31-Mar-1990 #text_change
ACCESSIONS 508059
REFERENCE 508059
#authors Chien, C.A.; Borroto, K.; Kamalay, J.A.; Dure III, L.
#journal Plant Mol. Biol. (1987) 9:533-546
#title Developmental biochemistry of cottonseed embryogenesis and germination. XIX. Sequences and genomic organization of the alpha globulin (vicilin) genes of cottonseed.
#accession 508059
#status not compared with conceptual translation
#molecule_type DNA
#residues 1-509 #label CHL
CLASSIFICATION #superfamily glycylin
SUMMARY #length 509 #checksum 9476
Query Match 50.28: Score 1910; DB 2; Length 509;
Best Local Similarity 56.98; Pred. No. 0.00e+00;
Matches 261; Conservative 95; Mismatches 92; Indels 11; Gaps 10;
Db 1 DPOARRVODCRHQOE--ERLRPHCEQSCREY--EKQOQOQPD--KQFKECOQRCOWOE 133
1 DPOARRVODCRHQOE--ERLRPHCEQSCREY--EKQOQOQPD--KQFKECOQRCOWOE 133
OY 37 DPOQVQEQOCORCESEATEEREEOQEOECERCEYKQEQOEEELORQVQOCGRQCO--EQ 95
37 DPOQVQEQOCORCESEATEEREEOQEOECERCEYKQEQOEEELORQVQOCGRQCO--EQ 95
Db 55 OPPER--KOQCVRECRYOYEDPMKGERENKMPREEEEDDEQOQRNNPYFH--RRSQ 191
55 OPPER--KOQCVRECRYOYEDPMKGERENKMPREEEEDDEQOQRNNPYFH--RRSQ 191
OY 97 COGQRCQOCCKCKCHQYKEOE--RGEHEN--YHNHKNRSEEEGQQRNNPYFPKRSPQ 153
97 COGQRCQOCCKCKCHQYKEOE--RGEHEN--YHNHKNRSEEEGQQRNNPYFPKRSPQ 153
Db 112 EPRFEHGNFVLOFPAKHLRGINFERIALILEANPTFVLPHHCDAEITYVYVINGSG 251
112 EPRFEHGNFVLOFPAKHLRGINFERIALILEANPTFVLPHHCDAEITYVYVINGSG 251
OY 156 TFRDEGCFKILQFPAKSPPLKGINDRFLAMFENPTFLPHHCDAEITYVYVINGSG 213
156 TFRDEGCFKILQFPAKSPPLKGINDRFLAMFENPTFLPHHCDAEITYVYVINGSG 213
Db 172 TITFTTHEKESYNVPGVYVIRIPAGSTYVLANODNREKLTIAVLRPVNPGOFQKFFP 311
172 TITFTTHEKESYNVPGVYVIRIPAGSTYVLANODNREKLTIAVLRPVNPGOFQKFFP 311
OY 216 TITFTTHEKESYNVPGVYVIRIPAGSTYVLANODNREKLTIAVLRPVNPGOFQKFFP 273
216 TITFTTHEKESYNVPGVYVIRIPAGSTYVLANODNREKLTIAVLRPVNPGOFQKFFP 273
Db 232 SORPOSTGALFRELLEAVNTRESEOLDELPGRQSRHROGO--GMFRKASQOIRAL 370
232 SORPOSTGALFRELLEAVNTRESEOLDELPGRQSRHROGO--GMFRKASQOIRAL 370
OY 276 NKPSTYGAFTVYVTEGNGEHCPCPHLSROSSD--SREEE--EOEEVEVRSQY 488
276 NKPSTYGAFTVYVTEGNGEHCPCPHLSROSSD--SREEE--EOEEVEVRSQY 488
Db 291 EATSPREKSGERFAFNLLSOTPRYSNOKRFEACPRNFQOOLREVDSVAFELINKG 430
291 EATSPREKSGERFAFNLLSOTPRYSNOKRFEACPRNFQOOLREVDSVAFELINKG 430
OY 336 SQGATSPKGSSEGTAFNLLSOTPRYSNOKRFEACPRNFQOOLREVDSVAFELINKG 392
336 SQGATSPKGSSEGTAFNLLSOTPRYSNOKRFEACPRNFQOOLREVDSVAFELINKG 392
Db 351 VPHYSKATFVYVTEGNGEHCPCPHLSROSSD--SREEE--EOEEVEVRSQY 488
351 VPHYSKATFVYVTEGNGEHCPCPHLSROSSD--SREEE--EOEEVEVRSQY 488
OY 396 VPHYSKATFVYVTEGNGEHCPCPHLSROSSD--SREEE--EOEEVEVRSQY 452
396 VPHYSKATFVYVTEGNGEHCPCPHLSROSSD--SREEE--EOEEVEVRSQY 452
Db 411 RSRQSLRILFRELLEAVNTRESEOLDELPGRQSRHROGO--GMFRKASQOIRAL 370
411 RSRQSLRILFRELLEAVNTRESEOLDELPGRQSRHROGO--GMFRKASQOIRAL 370
OY 456 KAPLSPGDVAVAPAGHATVFFASKDQPLNVAAGL--MNONNORITLAOR 503
456 KAPLSPGDVAVAPAGHATVFFASKDQPLNVAAGL--MNONNORITLAOR 503

(TM)

Distribution rights by Oxford University Press

protein database search, using Smith-Waterman

14.761 Million ce

generated.

~~(81-140) from US09331631 pep (3 of 3)~~

~~TABOYOCGBCOE000GCR.....HENYBNHKKNRS~~

.....

2AM 150
11

131253 seqs, 1295664 / residues

isting first 45 summaries

1:5A COMB 2:5B_COMB 3:PCT_COMB 4:backfiles

mean 23.260; variance 59.417; scale 0.000

ved by analysis of the total score distribution

SUMMARIES

Reply	Length	DB	ID	Description
0.0	566	1	US-07-955-	Sequence 2, Appl.
0.0	566	1	US-07-955-	Sequence 22, Appl
3.1	587	1	US-07-955-	Sequence 23, Appl
1.2	1898	2	US-08-800-	Sequence 94, Appl
1.2	1898	1	US-08-056-	Sequence 94, Appl
0.2	303	1	US-08-109-	Sequence 2, Appl
0.2	303	2	US-08-460-	Sequence 2, Appl
0.2	303	1	US-08-459-	Sequence 2, Appl
0.8	1162	2	US-08-728-	Sequence 2, Appl
8.8	788	2	US-08-918-	Sequence 4, Appl
8.7	34	2	US-08-822-	Sequence 1, Appl
7.8	361	1	US-08-115-	Sequence 4, Appl
7.6	176	2	US-08-889-	Sequence 3, Appl
7.0	303	1	US-08-185-	Sequence 5, Appl
7.0	344	4	5210183-2	Patent No. 52101
7.0	360	2	US-08-531-	Sequence 2, Appl
7.0	404	1	US-08-453-	Sequence 2, Appl
7.0	404	1	US-08-452-	Sequence 2, Appl
7.0	491	2	US-07-942-	Sequence 24, App
7.0	605	1	US-08-935-	Sequence 4, Appl
7.0	683	4	5210183-3	Patent No. 52101
7.0	737	1	US-08-185-	Sequence 4, Appl
7.0	737	1	US-08-185-	Sequence 2, Appl

24	79	16.8	559	1	US-08-313-	Sequence 14,	Applicat1	1.4ie+01
25	78	16.6	76	4	5482709-10	Patent No. 5482709,	Applicat1	1.69e+01
26	78	16.6	76	4	5273901-11	Patent No. 5273901,	Applicat1	1.88e+01
27	78	16.6	432	2	US-08-933-	Sequence 47,	Applicat1	1.69e+01
28	77	16.3	76	2	US-08-326-	Sequence 55,	Applicat1	2.02e+01
29	77	16.3	76	2	US-08-460-	Sequence 55,	Applicat1	2.02e+01
30	77	16.3	77	2	US-08-634-	Sequence 55,	Applicat1	2.02e+01
31	77	16.3	77	2	US-08-461-	Sequence 55,	Applicat1	2.02e+01
32	77	16.3	77	2	US-08-466-	Sequence 55,	Applicat1	2.02e+01
33	77	16.3	77	2	US-08-465-	Sequence 55,	Applicat1	2.02e+01
34	77	16.3	77	2	US-08-468-	Sequence 55,	Applicat1	2.02e+01
35	77	16.3	428	1	US-08-190-	Sequence 29,	Applicat1	2.42e+01
36	76	16.1	678	3	PCR-US93-0	Sequence 3,	Applicatio	2.42e+01
37	76	16.1	1045	2	US-08-451-	Sequence 6,	Applicatio	2.89e+01
38	75	15.9	393	2	US-08-467-	Sequence 9,	Applicatio	2.89e+01
39	75	15.9	816	2	US-08-267-	Sequence 4,	Applicatio	2.89e+01
40	75	15.9	1311	3	US-08-540-	Sequence 4,	Applicatio	2.89e+01
41	75	15.9	1311	3	PCR-US95-1	Sequence 4,	Applicatio	2.89e+01
42	75	15.9	2703	1	US-08-185-	Sequence 19,	Applicat1	3.46e+01
43	74	15.7	78	2	US-08-634-	Sequence 53,	Applicat1	3.46e+01
44	74	15.7	162	2	US-08-633-	Sequence 63,	Applicat1	3.46e+01
45	74	15.7	162	2	US-08-461-	Sequence 63,	Applicat1	3.46e+01

ALIGNMENTS

```

LT 1 STANDARD: 566 AA
US-07-955-905A-2
xxxxxx
Sequence 2, Application US/07955905A

```

Sequence 2, Application US/07955905A

Patent No. 5770433
GENERAL INFORMATION

APPLICANT: RECOMBINANT 47 AND 31 KD COCOA PROTEINS AND
TITLE OF INVENTION:

TITLE OF INVENTION:	PRE
NUMBER OF SEQUENCES:	28

COMPUTER READABLE FORM:

MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/955,905A

FILING DATE: 21-JAN-1993
CLASSIFICATION: A35

CLASSIFICATION: TOP SECRET
INFORMATION FOR SEQ ID NO: 1
SECURITY CHARACTERISTICS:

SEQUENCE CHARACTERISTICS:
LENGTH: 566 amino acids

TYPE: amino acid
TOPOLOGY: linear

MOLECULE TYPE: protein
SEQUENCE 566 AA; 65613 MW; 1529150 CN;

100.08: Score 471: DB 1: Length 566:

1st Local Similarity	100.00;	Pred. No. 4 67e-36;	Indels	0;	Gaps	0
Conservative	0;	Mismatches	0;			

[illegible][illegible]

81 LQRRYRCRCGEQDQDGRERQDQCKKCEQIKVEQENGENENINNNNNNNNSLEEDVNA

US-07-955-905A-22 STANDARD: PRT; 566 AA.

XXXXXX

M P R S C H (™)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

Mprsch_pp protein - protein database search, using Smith-Waterman algorithm
Run on: Sat May 13 12:12:57 2000: Maspar time 59.01 Seconds
7.246 Million cell updates/sec

Tabular output not generated.

Title: >US-09-331-631-23
Description: 287
Perfect Score: 1 RSGRGECCRQCLRRHGGCPWMEQCMRCRRG 33
Sequence:

Scoring table: PAM 150
Gap 11

Searched: 131253 seqs, 12956647 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: a-issued
1:5A_COMB 2:5B_COMB 3:PCT_COMB 4:backfiles1

Statistics: Mean 22.066; Variance 101.322; scale 0.218

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description	Pred. No.
1	287	100.0	34	2	US-08-822-Sequence 1, Applicatio	3.78e+17
2	87	30.3	188	2	US-08-569-Sequence 11, Applicati	3.85e+00
3	87	30.3	188	1	US-08-469-Sequence 11, Applicati	3.85e+00
4	87	30.3	188	2	US-08-609-Sequence 11, Applicati	4.59e+00
5	86	30.0	566	1	US-07-955-Sequence 2, Applicatio	4.59e+00
6	86	30.0	566	1	US-07-955-Sequence 22, Applicati	1.56e+01
7	79	27.5	587	1	US-07-955-Sequence 23, Applicati	3.67e+01
8	74	25.8	55	2	US-08-609-Sequence 3, Applicatio	3.67e+01
9	74	25.8	55	2	US-08-469-Sequence 3, Applicatio	3.67e+01
10	74	25.8	55	1	US-08-469-Sequence 3, Applicatio	3.67e+01
11	74	25.8	188	2	US-08-569-Sequence 5, Applicatio	3.67e+01
12	74	25.8	188	2	US-08-469-Sequence 5, Applicatio	3.67e+01
13	74	25.8	188	1	US-08-469-Sequence 5, Applicatio	3.67e+01
14	74	25.8	188	2	US-08-609-Sequence 7, Applicatio	3.67e+01
15	74	25.8	188	2	US-08-469-Sequence 7, Applicatio	3.67e+01
16	74	25.8	188	2	US-08-569-Sequence 7, Applicatio	3.67e+01
17	74	25.8	188	2	US-08-469-Sequence 7, Applicatio	3.67e+01
18	74	25.8	188	2	US-08-569-Sequence 7, Applicatio	3.67e+01
19	74	25.8	188	2	US-08-469-Sequence 7, Applicatio	3.67e+01
20	74	25.8	188	2	US-08-569-Sequence 7, Applicatio	3.67e+01
21	74	25.8	188	2	US-08-469-Sequence 7, Applicatio	3.67e+01
22	74	25.8	188	2	US-08-569-Sequence 7, Applicatio	3.67e+01
23	74	25.8	188	2	US-08-469-Sequence 7, Applicatio	3.67e+01

24	69	24.0	1101	2	US-08-469-Sequence 96, Applicati	8.51e+01
25	69	24.0	1118	1	US-08-278-Sequence 2, Applicatio	8.51e+01
26	69	24.0	1118	1	US-08-838-Sequence 2, Applicatio	8.51e+01
27	69	24.0	1118	1	US-07-934-Sequence 2, Applicatio	8.51e+01
28	69	24.0	1122	1	US-08-838-Sequence 6, Applicatio	8.51e+01
29	69	24.0	1122	1	US-08-278-Sequence 6, Applicatio	8.51e+01
30	69	24.0	1122	1	US-08-505-Sequence 7, Applicatio	1.01e+02
31	69	24.0	1122	1	US-09-018-Sequence 7, Applicatio	1.01e+02
32	69	24.0	1122	1	US-08-656-Sequence 25, Applicatio	1.01e+02
33	69	24.0	1122	1	US-08-670-Sequence 6, Applicatio	1.01e+02
34	69	24.0	1122	1	US-08-477-Sequence 1, Applicatio	1.01e+02
35	69	24.0	1122	1	US-08-484-Sequence 1, Applicatio	1.01e+02
36	69	24.0	1122	1	US-08-477-Sequence 2, Applicatio	1.01e+02
37	69	24.0	1122	1	US-08-464-Sequence 2, Applicatio	1.01e+02
38	69	24.0	1122	1	US-08-243-Sequence 3, Applicatio	1.01e+02
39	69	24.0	1122	1	US-08-477-Sequence 3, Applicatio	1.01e+02
40	69	24.0	1122	1	US-08-477-Sequence 3, Applicatio	1.01e+02
41	69	24.0	1122	1	US-08-477-Sequence 3, Applicatio	1.01e+02
42	69	24.0	1122	1	US-08-477-Sequence 3, Applicatio	1.01e+02
43	69	24.0	1122	1	US-08-477-Sequence 3, Applicatio	1.01e+02
44	69	24.0	1122	1	US-08-477-Sequence 3, Applicatio	1.01e+02
45	69	24.0	1122	1	US-08-477-Sequence 3, Applicatio	1.01e+02

ALIGNMENTS

RESULT 1 STANDARD: PRT: 34 AA.
ID US-08-822-561-1
AC xxxxxx
XX
DT
XX Sequence 1, Application US/08822561
DE
XX Sequence 1, Application US/08822561
CC Patent No. 590187
CC GENERAL INFORMATION:
CC APPLICANT: Duwick, John P.
CC APPLICANT: Rood, Tracy A.
CC APPLICANT: Rao, Aravula G.
CC TITLE OF INVENTION: ANTIMICROBIAL PEPTIDES AND PLANT DISEASE
CC TITLE OF INVENTION: RESISTANCE BASED THEREON
CC NUMBER OF SEQUENCES: 3
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: PIONEER HI-BRED INTERNATIONAL, INC.
CC STREET: 700 Capital Square, 400 Locust Street
CC CITY: Des Moines
CC STATE: Iowa
CC COUNTRY: USA
CC ZIP: 50309
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: Patentin Release #1.0, Version #1.30
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/822,561
CC FILING DATE: 19-MAR-1997
CC CLASSIFICATION: 800
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 08/442,910
CC FILING DATE: 17-MAY-1995
CC APPLICATION NUMBER: US 07/536,127
CC FILING DATE: 05-JUN-1990
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Yates, Michael J.
CC REGISTRATION NUMBER: 36,063
CC REFERENCE/DOCKET NUMBER: 0124C
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (515) 248-4800
CC TELEFAX: (515) 248-4844
CC INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:
LENGTH: 34 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE 34 AA: 4188 MW: 3921 CN:
Query Match 100.0%; Score 287; DB 2; Length 34;
Best Local Similarity 100.0%; Pred. No. 3,78e-17;
Matches 33; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
DB 1 RSGRGCRROCLRRHGGPMTQECMRRCRRG 33
OY 1 RSGRGCRROCLRRHGGPMTQECMRRCRRG 33
RESULT 2
ID US-08-569-063C-11 STANDARD: PRT: 188 AA.
AC xxxxxx
DT
XX
DE Sequence 11, Application US/08569063C
XX
CC Sequence 11, Application US/08569063C
CC Patent No. 5928939
CC GENERAL INFORMATION:
CC APPLICANT: ERIKSSON, Ulf
CC APPLICANT: OLOFSSON, Birgitta
CC APPLICANT: ALLTALO, Kari
CC APPLICANT: PAJUSOLA, Katri
CC TITLE OF INVENTION: VASCULAR ENDOTHELIAL GROWTH FACTOR-B AND
CC TITLE OF INVENTION: DNA CODING THEREFOR
CC NUMBER OF SEQUENCES: 23
CC CORRESPONDENCE ADDRESSES:
CC ADDRESSEE: Evenson, McKeown, Edwards & Lenahan, P.L.L.C.
CC STREET: 1200 G Street, N.W., Suite 700
CC CITY: Washington
CC STATE: DC
CC COUNTRY: USA
CC ZIP: 20005
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: floppy disk
CC OPERATING SYSTEM: IBM PC compatible
CC SOFTWARE: Patentin Release #1.0, Version #1.25
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/569, 063C
CC FILING DATE: 06-DEC-1995
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 08/469, 427
CC FILING DATE: 06-JUN-1995
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 08/397, 651
CC FILING DATE: 01-MAR-1995
CC ATTORNEY/AGENT INFORMATION:
CC NAME: EVANS, Joseph D
CC REGISTRATION NUMBER: 26,269
CC REFERENCE/DOCKET NUMBER: 1064/41979CP3
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (202) 628-8800
CC TELEFAX: (202) 628-8844
CC INFORMATION FOR SEQ ID NO: 11:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 188 amino acids
CC TYPE: amino acid
CC STRANDEDNESS: single
CC TOPOLOGY: linear
CC MOLECULE TYPE: protein
CC HYPOTHETICAL: NO
CC ORIGINAL SOURCE:
CC SIGNAL SOURCE:

TISSUE TYPE: human fibrosarcoma
SEQUENCE 188 AA: 21261 MW: 161002 CN:
Query Match 30.3%; Score 87; DB 2; Length 188;
Best Local Similarity 40.6%; Pred. No. 3,85e+00;
Matches 13; Conservative 7; Mismatches 9; Indels 3; Gaps 3;
DB 137 SPRPLCPR-CTOHNO-RP-DPRTCRCRCRRS 165
OY 2 SGRGCRROCLRRHGGPMTQECMRRCRRG 33
RESULT 3
ID US-08-469-427A-11 STANDARD: PRT: 188 AA.
AC xxxxxx
DT
XX
DE Sequence 11, Application US/08469427A
XX
CC Sequence 11, Application US/08469427A
CC Patent No. 5607918
CC GENERAL INFORMATION:
CC APPLICANT: ERIKSSON, Ulf
CC APPLICANT: OLOFSSON, Birgitta
CC APPLICANT: ALLTALO, Kari
CC APPLICANT: PAJUSOLA, Katri
CC TITLE OF INVENTION: VASCULAR ENDOTHELIAL GROWTH FACTOR-B AND
CC TITLE OF INVENTION: DNA CODING THEREFOR
CC NUMBER OF SEQUENCES: 17
CC CORRESPONDENCE ADDRESSES:
CC ADDRESSEE: Evenson, McKeown, Edwards & Lenahan
CC STREET: 1200 G Street, N.W., Suite 700
CC CITY: Washington
CC STATE: DC
CC ZIP: 20005
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: floppy disk
CC OPERATING SYSTEM: IBM PC compatible
CC SOFTWARE: Patentin Release #1.0, Version #1.25
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/469, 427A
CC FILING DATE: 06-JUN-1995
CC CLASSIFICATION: 435
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 08/397, 651
CC FILING DATE: 01-MAR-1995
CC ATTORNEY/AGENT INFORMATION:
CC NAME: EVANS, Joseph D
CC REGISTRATION NUMBER: 26,269
CC REFERENCE/DOCKET NUMBER: 41979CP2
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (202) 628-8800
CC TELEFAX: (202) 628-8844
CC INFORMATION FOR SEQ ID NO: 11:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 188 amino acids
CC TYPE: amino acid
CC STRANDEDNESS: single
CC TOPOLOGY: linear
CC MOLECULE TYPE: protein
CC HYPOTHETICAL: NO
CC ORIGINAL SOURCE:
CC TISSUE TYPE: human fibrosarcoma
CC SEQUENCE 188 AA: 21261 MW: 161002 CN:
Query Match 30.3%; Score 87; DB 1; Length 188;
Best Local Similarity 40.6%; Pred. No. 3,85e+00;
Matches 13; Conservative 7; Mismatches 9; Indels 3; Gaps 3;
DB 137 SPRPLCPR-CTOHNO-RP-DPRTCRCRCRRS 165

WISER (TM)

Release 3.1A John F. Collins, BioComputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MParch_dp protein - protein database search, using Smith-Waterman algorithm
Run on: Sat May 13 09:00:19 2000; Maspar time 78.64 Seconds
8.907 Million cell updates/sec

Tabular output not generated.

Title: >US-09-331-631-25
Description: (1-23) from US0931631.pep
Perfect Score: 177
Sequence: 1 MMRARFPLLLGLVFLASVSF 23

Scoring table: PAM 150
Gap 11

Searched: 83857 seqs, 30454973 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: swiss-prot38
1:swissprot

Statistics: Mean 33.887; Variance 71.622; scale 0.473

Pred: No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description	Pred. No.
1	177	100.0	605	1	GLCA_SOYBN	BETA-CONGLYCININ, ALPH	2,46e-15
2	170	96.0	639	1	GLCX_SOYBN	BETA-CONGLYCININ, ALPH	3,71e-14
3	138	78.0	421	1	PHSB_PHAVU	PHASEOLIN, BETA-TYPE P	6,39e-09
4	138	78.0	436	1	PHSA_PHAVU	PHASEOLIN, ALPHA-TYPE	6,39e-09
5	136	76.8	439	1	GLCB_SOYBN	BETA-CONGLYCININ, BETA	1,33e-08
6	120	67.8	571	1	CVCB_PEA	CONVICILIN PRECURSOR	4,08e-06
7	114	64.4	386	1	CVCB_PEA	CONVICILIN PRECURSOR	3,28e-05
8	111	62.7	445	1	CANA_CANLN	CANAVALIN PRECURSOR	9,18e-05
9	111	62.7	445	1	CANA_CANLN	CANAVALIN PRECURSOR	9,18e-05
10	105	59.3	459	1	VCLC_PEA	VICILIN PRECURSOR	6,96e-04
11	94	53.1	463	1	VCLC_PEA	VICILIN PRECURSOR	2,52e-02
12	84	47.5	547	1	CP78_MAIZE	CYCLOCHROME P450 78A1	5,57e-01
13	82	46.3	659	1	VST2_HEVBE	STRUCTURAL PROTEIN 2 P	1,36e+00
14	81	45.8	502	1	V752_BORBU	HYPOTHELICAL PROTEIN B	1,36e+00
15	81	45.8	614	1	AH12_ARAHY	ALLERGEN ARA H 1, CLON	1,36e+00
16	81	45.8	626	1	AH12_ARAHY	ALLERGEN ARA H 1, CLON	1,36e+00
17	81	45.8	909	1	LDL1_XENLA	LOW-DENSITY LIPOPROTEI	1,83e+00
18	80	45.2	141	1	SCRC_RAT	SECRETIN RECEPTOR PREC	1,44e+00
19	80	45.2	449	1	SCRC_RAT	SECRETIN RECEPTOR PREC	1,44e+00
20	79	44.6	67	1	YPHE_BACSU	HYPOTHELICAL 7.5 KD PR	2,44e+00
21	79	44.6	405	1	WCAD_ECOLI	PUTATIVE COLANIC ACID	4,35e+00
22	77	43.5	124	1	VBF3_YEAST	VERY HYPOTHELICAL 13.2	4,35e+00
23	77	43.5	660	1	VST2_HEVBE	STRUCTURAL PROTEIN 2 P	4,35e+00

SEQ ID NO: 25

RESULT	ID	GLCA_SOYBN	STANDARD	PRT	605 AA.	ALIGNMENTS
24	77	43.5	660	1	VST2_HEVBE	STRUCTURAL PROTEIN 2 P
25	77	43.5	660	1	VST2_HEVBE	STRUCTURAL PROTEIN 2 P
26	76	42.9	213	1	IT14_HUMAN	IMMUNOGLOBULIN-RELATED
27	76	42.9	214	1	GLP1_PHANI	GERMIN-LIKE PROTEIN PR
28	76	42.9	336	1	MRA1_CHLTR	PROSPERO-N-ACETYLIMBRO
29	75	42.4	769	1	ITB2_BOVIN	CELL SURFACE ADHESION
30	75	42.4	979	1	AMD_MOUSE	PEPTIDYL-GLYCINE ALPHA
31	74	41.8	173	1	PRL_MEDTR	PATHOGENESIS-RELATED P
32	74	41.8	412	1	GLUP_BRUBA	GLUCOSE/GALACTOSE TRAN
33	74	41.8	603	1	NUSM_GORGO	NADH-UBIQUINONE OXIDOR
34	73	41.2	106	1	PSIF_ECOLI	PHOSPHATE STARVATION-I
35	73	41.2	202	1	NUOJ_RHOCA	NADH DEHYDROGENASE I C
36	73	41.2	393	1	YBBD_BACSU	HYPOTHELICAL 42.7 KD L
37	73	41.2	642	1	XKXZ_CLOTH	ENDO-1,4-BETA-XYLANASE
38	73	41.2	837	1	LDL2_XENLA	LOW-DENSITY LIPOPROTEI
39	73	41.2	892	1	TORS_ECOLI	SENSOR PROTEIN TORS
40	73	41.2	904	1	POLG_HCVJ6	GENOME POLYPROTEIN
41	73	41.2	3033	1	LCA_CAVPO	ALPHA-LACTALBUMIN
42	72	40.7	142	1	LPPB_HAEIN	OUTER MEMBRANE ANTIGEN
43	72	40.7	405	1	LPPB_HAEIN	OUTER MEMBRANE ANTIGEN
44	72	40.7	669	1	FP51_YEAST	GLYCEROL UPTAKE/EFFLUX
45	72	40.7	690	1	PERO_DROME	PEROXIDASE PRECURSOR

1
GLCA_SOYBN
BETA-CONGLYCININ, ALPH
2,46e-15
3,71e-14
6,39e-09
6,39e-09
1,33e-08
4,08e-06
3,28e-05
9,18e-05
6,96e-04
2,52e-02
5,57e-01
1,36e+00
1,36e+00
1,36e+00
1,36e+00
1,83e+00
1,44e+00
1,44e+00
2,44e+00
4,35e+00
4,35e+00

ALIGNMENTS

1
GLCA_SOYBN
STANDARD
PRT
605 AA.

1
GLCA_SOYBN
BETA-CONGLYCININ, ALPH
2,46e-15
3,71e-14
6,39e-09
6,39e-09
1,33e-08
4,08e-06
3,28e-05
9,18e-05
6,96e-04
2,52e-02
5,57e-01
1,36e+00
1,36e+00
1,36e+00
1,36e+00
1,83e+00
1,44e+00
1,44e+00
2,44e+00
4,35e+00
4,35e+00

1
GLCA_SOYBN
BETA-CONGLYCININ, ALPH
2,46e-15
3,71e-14
6,39e-09
6,39e-09
1,33e-08
4,08e-06
3,28e-05
9,18e-05
6,96e-04
2,52e-02
5,57e-01
1,36e+00
1,36e+00
1,36e+00
1,36e+00
1,83e+00
1,44e+00
1,44e+00
2,44e+00
4,35e+00
4,35e+00

1
GLCA_SOYBN
BETA-CONGLYCININ, ALPH
2,46e-15
3,71e-14
6,39e-09
6,39e-09
1,33e-08
4,08e-06
3,28e-05
9,18e-05
6,96e-04
2,52e-02
5,57e-01
1,36e+00
1,36e+00
1,36e+00
1,36e+00
1,83e+00
1,44e+00
1,44e+00
2,44e+00
4,35e+00
4,35e+00

1
GLCA_SOYBN
BETA-CONGLYCININ, ALPH
2,46e-15
3,71e-14
6,39e-09
6,39e-09
1,33e-08
4,08e-06
3,28e-05
9,18e-05
6,96e-04
2,52e-02
5,57e-01
1,36e+00
1,36e+00
1,36e+00
1,36e+00
1,83e+00
1,44e+00
1,44e+00
2,44e+00
4,35e+00
4,35e+00

1
GLCA_SOYBN
BETA-CONGLYCININ, ALPH
2,46e-15
3,71e-14
6,39e-09
6,39e-09
1,33e-08
4,08e-06
3,28e-05
9,18e-05
6,96e-04
2,52e-02
5,57e-01
1,36e+00
1,36e+00
1,36e+00
1,36e+00
1,83e+00
1,44e+00
1,44e+00
2,44e+00
4,35e+00
4,35e+00

1
GLCA_SOYBN
BETA-CONGLYCININ, ALPH
2,46e-15
3,71e-14
6,39e-09
6,39e-09
1,33e-08
4,08e-06
3,28e-05
9,18e-05
6,96e-04
2,52e-02
5,57e-01
1,36e+00
1,36e+00
1,36e+00
1,36e+00
1,83e+00
1,44e+00
1,44e+00
2,44e+00
4,35e+00
4,35e+00

1
GLCA_SOYBN
BETA-CONGLYCININ, ALPH
2,46e-15
3,71e-14
6,39e-09
6,39e-09
1,33e-08
4,08e-06
3,28e-05
9,18e-05
6,96e-04
2,52e-02
5,57e-01
1,36e+00
1,36e+00
1,36e+00
1,36e+00
1,83e+00
1,44e+00
1,44e+00
2,44e+00
4,35e+00
4,35e+00

1
GLCA_SOYBN
BETA-CONGLYCININ, ALPH
2,46e-15
3,71e-14
6,39e-09
6,39e-09
1,33e-08
4,08e-06
3,28e-05
9,18e-05
6,96e-04
2,52e-02
5,57e-01
1,36e+00
1,36e+00
1,36e+00
1,36e+00
1,83e+00
1,44e+00
1,44e+00
2,44e+00
4,35e+00
4,35e+00

1
GLCA_SOYBN
BETA-CONGLYCININ, ALPH
2,46e-15
3,71e-14
6,39e-09
6,39e-09
1,33e-08
4,08e-06
3,28e-05
9,18e-05
6,96e-04
2,52e-02
5,57e-01
1,36e+00
1,36e+00
1,36e+00
1,36e+00
1,83e+00
1,44e+00
1,44e+00
2,44e+00
4,35e+00
4,35e+00

1
GLCA_SOYBN
BETA-CONGLYCININ, ALPH
2,46e-15
3,71e-14
6,39e-09
6,39e-09
1,33e-08
4,08e-06
3,28e-05
9,18e-05
6,96e-04
2,52e-02
5,57e-01
1,36e+00
1,36e+00
1,36e+00
1,36e+00
1,83e+00
1,44e+00
1,44e+00
2,44e+00
4,35e+00
4,35e+00

1
GLCA_SOYBN
BETA-CONGLYCININ, ALPH
2,46e-15
3,71e-14
6,39e-09
6,39e-09
1,33e-08
4,08e-06
3,28e-05
9,18e-05
6,96e-04
2,52e-02
5,57e-01
1,36e+00
1,36e+00
1,36e+00
1,36e+00
1,83e+00
1,44e+00
1,44e+00
2,44e+00
4,35e+00
4,35e+00

1
GLCA_SOYBN
BETA-CONGLYCININ, ALPH
2,46e-15
3,71e-14
6,39e-09
6,39e-09
1,33e-08
4,08e-06
3,28e-05
9,18e-05
6,96e-04
2,52e-02
5,57e-01
1,36e+00
1,36e+00
1,36e+00
1,36e+00
1,83e+00
1,44e+00
1,44e+00
2,44e+00
4,35e+00
4,35e+00

1
GLCA_SOYBN
BETA-CONGLYCININ, ALPH
2,46e-15
3,71e-14
6,39e-09
6,39e-09
1,33e-08
4,08e-06
3,28e-05
9,18e-05
6,96e-04
2,52e-02
5,57e-01
1,36e+00
1,36e+00
1,36e+00
1,36e+00
1,83e+00
1,44e+00
1,44e+00
2,44e+00
4,35e+00
4,35e+00

1
GLCA_SOYBN
BETA-CONGLYCININ, ALPH
2,46e-15
3,71e-14
6,39e-09
6,39e-09
1,33e-08
4,08e-06
3,28e-05
9,18e-05
6,96e-04
2,52e-02
5,57e-01
1,36e+00
1,36e+00
1,36e+00
1,36e+00
1,83e+00
1,44e+00
1,44e+00
2,44e+00
4,35e+00
4,35e+00

1
GLCA_SOYBN
BETA-CONGLYCININ, ALPH
2,46e-15
3,71e-14
6,39e-09
6,39e-09
1,33e-08
4,08e-06
3,28e-05
9,18e-05
6,96e-04
2,52e-02
5,57e-01
1,36e+00
1,36e+00
1,36e+00
1,36e+00
1,83e+00
1,44e+00
1,44e+00
2,44e+00
4,35e+00
4,35e+00

1
GLCA_SOYBN
BETA-CONGLYCININ, ALPH
2,46e-15
3,71e-14
6,39e-09
6,39e-09
1,33e-08
4,08e-06
3,28e-05
9,18e-05
6,96e-04
2,52e-02
5,57e-01
1,36e+00
1,36e+00
1,36e+00
1,36e+00
1,83e+00
1,44e+00
1,44e+00
2,44e+00
4,35e+00
4,35e+00

1
GLCA_SOYBN
BETA-CONGLYCININ, ALPH
2,46e-15
3,71e-14
6,39e-09
6,39e-09
1,33e-08
4,08e-06
3,28e-05
9,18e-05
6,96e-04
2,52e-02
5,57e-01
1,36e+00
1,36e+00
1,36e+00
1,36e+00
1,83e+00
1,44e+00
1,44e+00
2,44e+00
4,35e+00
4,35e+00

1
GLCA_SOYBN
BETA-CONGLYCININ, ALPH
2,46e-15
3,71e-14
6,39e-09
6,39e-09
1,33e-08
4,08e-06
3,28e-05
9,18e-05
6,96e-04
2,52e-02
5,57e-01
1,36e+00
1,36e+00
1,36e+00
1,36e+00
1,83e+00
1,44e+00
1,44e+00
2,44e+00
4,35e+00
4,35e+00

1
GLCA_SOYBN
BETA-CONGLYCININ, ALPH
2,46e-15
3,71e-14
6,39e-09
6,39e-09
1,33e-08
4,08e-06
3,28e-05
9,18e-05
6,96e-04
2,52e-02
5,57e-01
1,36e+00
1,36e+00
1,36e+00
1,36e+00
1,83e+00
1,44e+00
1,44e+00
2,44e+00
4,35e+00
4,35e+00

1
GLCA_SOYBN
BETA-CONGLYCININ, ALPH
2,46e-15
3,71e-14
6,39e-09
6,39e-09
1,33e-08
4,08e-06
3,28e-05
9,18e-05
6,96e-04
2,52e-02
5,57e-01
1,36e+00
1,36e+00
1,36e+00
1,36e+00
1,83e+00
1,44e+00
1,44e+00
2,44e+00
4,35e+00
4,35e+00

1
GLCA_SOYBN
BETA-CONGLYCININ, ALPH
2,46e-15
3,71e-14
6,39e-09
6,39e-09
1,33e-08
4,08e-06
3,28e-05
9,18e-05
6,96e-04
2,52e-02
5,57e-01
1,36e+00
1,36e+00
1,36e+00
1,36e+00
1,83e+00
1,44e+00
1,44e+00
2,44e+00
4,35e+00
4,35e+00

1
GLCA_SOYBN
BETA-CONGLYCININ, ALPH
2,46e-15
3,71e-14
6,39e-09
6,39e-09
1,33e-08
4,08e-06
3,28e-05
9,18e-05
6,96e-04
2,52e-02
5,57e-01
1,36e+00
1,36e+00
1,36e+00
1,36e+00
1,83e+00
1,44e+00
1,44e+00
2,44e+00
4,35e+00
4,35e+00

1
GLCA_SOYBN
BETA-CONGLYCININ, ALPH
2,46e-15
3,71e-14
6,39e-09
6,39e-09
1,33e-08
4,08e-06
3,28e-05
9,18e-05
6,96e-04
2,52e-02
5,57e-01
1,36e+00
1,36e+00
1,36e+00
1,36e+00
1,83e+00
1,44e+00
1,44e+00
2,44e+00
4,35e+00
4,35e+00

1
GLCA_SOYBN
BETA-CONGLYCININ, ALPH
2,46e-15
3,71e-14
6,39e-09
6,39e-09
1,33e-08
4,08e-06
3,28e-05
9,18e-05
6,96e-04
2,52e-02
5,57e-01
1,36e+00
1,36e+00
1,36e+00
1,36e+00
1,83e+00
1,44e+00
1,44e+00
2,44e+00
4,35e+00
4,35e+00

1
GLCA_SOYBN
BETA-CONGLYCININ, ALPH
2,46e-15
3,71e-14
6,39e-09
6,39e-09
1,33e-08
4,08e-06
3,28e-05
9,18e-05
6,96e-04
2,52e-02
5,57e-01
1,36e+00
1,36e+00
1,36e+00
1,36e+00
1,83e+00
1,44e+00
1,44e+00
2,44e+00
4,35e+00
4,35e+00

1
GLCA_SOYBN
BETA-CONGLYCININ, ALPH
2,46e-15
3,71e-14
6,39e-09
6,39e-09
1,33e-08
4,08e-06
3,28e-05
9,18e-05
6,96e-04
2,52e-02
5,57e-01
1,36e+00
1,36e+00
1,36e+00
1,36e+00
1,83e+00
1,44e+00
1,44e+00
2,44e+00
4,35e+00
4,35e+00

1
GLCA_SOYBN
BETA-CONGLYCININ, ALPH
2,46e-15
3,71e-14
6,39e-09
6,39e-09
1,33e-08
4,08e-06
3,28e-05
9,18e-05
6,96e-04
2,52e-02
5,57e-01
1,36e+00
1,36e+00
1,36e+00
1,36e+00
1,83e+00
1,44e+00
1,44e+00
2,44e+00
4,35e+00
4,35e+00

1
GLCA_SOYBN
BETA-CONGLYCININ, ALPH
2,46e-15
3,71e-14
6,39e-09
6,39e-09
1,33e-08
4,08e-06
3,28e-05
9,18e-05
6,96e-04
2,52e-02
5,57e-01
1,36e+00
1,36e+00
1,36e+00
1,36e+00
1,83e+00
1,44e+00
1,44e+00
2,44e+00
4,35e+00
4,35e+00

1
GLCA_SOYBN
BETA-CONGLYCININ, ALPH
2,46e-15
3,71e-14
6,39e-09
6,39e-09
1,33e-08
4,08e-06
3,28e-05
9,18e-05
6,96e-04
2,52e-02
5,57e-01
1,36e+00
1,36e+00
1,36e+00
1,36e+00
1,83e+00
1,44e+00
1,44e+00
2,44e+00
4,35e+00
4,35e+00

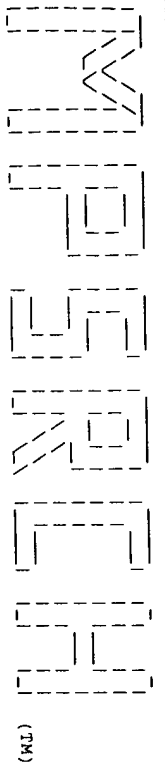
1
GLCA_SOYBN
BETA-CONGLYCININ, ALPH
2,46e-15
3,71e-14
6,39e-09
6,39e-09
1,33e-08
4,08e-06
3,28e-05
9,18e-05
6,96e-04
2,52e-02
5,57e-01
1,36e+00
1,36e+00
1,36e+00
1,36e+00
1,83e+00
1,44e+00
1,44e+00
2,44e+00
4,35e+00
4,35e+00

1
GLCA_SOYBN
BETA-CONGLYCININ, ALPH
2,46e-15
3,71e-14
6,39e-09
6,39e-09
1,33e-08
4,08e-06
3,28e-05
9,18e-05
6,96e-04
2,52e-02
5,57e-01
1,36e+00
1,36e+00
1,36e+00
1,36e+00
1,83e+00
1,44e+00
1,44e+00
2,44e+00
4,35e+00
4,35e+00

1
GLCA_SOYBN
BETA-CONGLYCININ, ALPH
2,46e-15
3,71e-14
6,39e-09
6,39e-09
1,33e-08
4,08e-06
3,28e-05
9,18e-05
6,96e-04
2,52e-02
5,57e-01
1,36e+00
1,36e+00
1,36e+00
1,36e+00
1,83e+00
1,44e+00
1,44e+00
2,44e+00
4,35e+00
4,35e+00

1
GLCA_SOYBN
BETA-CONGLYCININ, ALPH
2,46e-15
3,71e-14
6,39e-09
6,39e-09
1,33e-08
4,08e-06
3,28e-05
9,18e-05
6,96e-04
2,52e-02
5,57e-01
1,36e+00
1,36e+00
1,36e+00
1,36e+00
1,83e+00
1,44e+00
1,44e+00
2,44e+00
4,35e+00
4,35e+00

1
GLCA_SOYBN
BETA-CONGLYCININ, ALPH
2,46e-15
3,71e-14
6,39e-09
6,39e-09
1,33e-08
4,08e-06
3,28e-05
9,18e-05
6,96e-04
2,52e-02
5,57e-01
1,36e+00
1,36e+00
1,36e+00



Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

Search: protein - protein database search, using Smith-Waterman algorithm
Run on: Sat May 13 08:51:56 2000; MasPar time 6.17 Seconds
252.411 Million cell updates/sec

Tabular output not generated.

Title: >US-09-331-631-23
Description: (1-33) from 450331631.pep
Perfect Score: 1 RSGRGECCROCLRRHGECPWETOCMRCRRRG 33
Sequence: PAM 150
Scoring table: Gap 11

Searched: 142080 seqs, 47172406 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: p1r62
1:p1r1 2:p1r2 3:p1r3 4:p1r4

Statistics: Mean 30.542; Variance 60.292; Scale 0.507

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description	Pred. No.
1	287	100.0	33	2	A41822	antimicrobial peptide	7.27e-39
2	92	32.1	442	2	H72266	astB/chur-related pro	5.27e-03
3	91	31.7	122	2	C53234	globulin-10 - maize	7.50e-03
4	91	31.7	582	2	B53234	vicillin-like storage	7.50e-03
5	91	31.7	588	1	FWCNA8	alpha-globulin B prec	2.14e-02
6	88	30.7	509	2	S08059	alpha-globulin type B	4.27e-02
7	86	30.0	221	2	C34768	ORF2 protein - Ort vi	4.27e-02
8	86	30.0	566	2	S22477	vicillin precursor - c	6.02e-02
9	85	29.6	236	2	T01662	globulin-1 - maize (f	6.02e-02
10	85	29.6	540	2	T02258	globulin-1 - maize	6.02e-02
11	85	29.6	407	2	S21825	vicillin-like storage	6.02e-02
12	85	29.6	573	2	A53234	globulin-1S, GBLIS -	8.46e-02
13	84	29.3	52	2	I50722	neuron-glia adhesion	8.46e-02
14	82	28.6	527	2	A54985	BMP receptor precursor	1.66e-01
15	81	28.2	358	2	I50506	gene wnt8b protein -	2.33e-01
16	81	28.2	428	2	I51680	gene wnt8b protein -	3.24e-01
17	80	27.9	112	2	I51578	acu-8 protein - Neuro	4.52e-01
18	80	27.9	525	2	A36316	proctamine 2 - pig	4.52e-01
19	79	27.5	91	2	S21673	probable transposase	6.28e-01
20	79	27.5	344	2	E70950	omega-hordothionin -	6.28e-01
21	78	27.2	48	2	S68952	gamma-thionin precus	8.70e-01
22	78	27.2	82	2	S65779	protamine II-1 - pain	
23	77	26.8	45	2	B58208		

24	77	26.8	58	2	A58208	protamine II-1 - pain	8.70e-01
25	77	26.8	187	2	G64323	ribosomal protein S4	8.70e-01
26	77	26.8	605	2	S06398	alpha-globulin type A	8.70e-01
27	77	26.8	4976	2	T14165	peptide synthetase ho	1.20e+00
28	76	26.5	45	2	B69810	hypothetical protein	1.20e+00
29	76	26.5	47	2	E58208	protamine II-4 - pain	1.20e+00
30	76	26.5	51	2	B75130	lsu ribosomal protein	1.20e+00
31	76	26.5	51	2	C71001	probable ribosomal pr	1.20e+00
32	76	26.5	102	2	S33332	protamine P2 - pygmy	1.20e+00
33	76	26.5	102	2	S33333	protamine P2 - gotill	1.20e+00
34	76	26.5	102	2	S33331	protamine P2 - chimpa	1.20e+00
35	74	25.8	45	2	C58208	protamine II-2 - pain	2.29e+00
36	74	25.8	47	2	F58208	protamine II-5 - pain	2.29e+00
37	74	25.8	102	1	HS0UP2	sperm histone P2 prec	2.29e+00
38	74	25.8	188	2	JC4680	vascular endothelial	2.29e+00
39	74	25.8	1073	2	S56220	probable membrane pro	2.29e+00
40	74	25.8	1306	2	T13592	hypothetical protein	3.14e+00
41	73	25.4	58	2	S10755	protamine St2b - hors	3.14e+00
42	73	25.4	62	2	S10754	protamine St2a - hors	3.14e+00
43	73	25.4	92	2	S13132	protamine 2 precursor	3.14e+00
44	73	25.4	102	2	S33336	protamine P2 - rhesus	3.14e+00
45	73	25.4	4006	2	T09070	probable tenascin X -	3.14e+00

ALIGNMENTS

RESULT 1
ENTRY A41822 *type complete
TITLE antimicrobial peptide - maize
ORGANISM Zea mays
DATE 28-May-1992 *sequence-revision 28-May-1992 *text-change 30-Sep-1993

ACCESSIONS A41822
REFERENCE DAVICK, J.P., ROOD, T., RAO, A.G., MARSHAK, D.R.,
#authors submitted to the Protein Sequence Database, May 1992
#accession A41822
#status preliminary
#molecule-type protein
#residues 1-33 *label DUV

SUMMARY
#length 33 #molecular-weight 4131 #checksum 3118

Query Match 100.0%; Score 287; DB 2; Length 33;
Best Local Similarity 100.0%; Pred. No. 7.27e-39;
Matches 33; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 RSGRGECCROCLRRHGECPWETOCMRCRRRG 33
QY 1 RSGRGECCROCLRRHGECPWETOCMRCRRRG 33

RESULT 2
ENTRY H72266 *type complete
TITLE astB/chur-related protein - Thermotoga maritima (strain MSB8)
ORGANISM Thermotoga maritima
DATE 11-Jun-1999 *sequence-revision 11-Jun-1999 *text-change 11-Jun-1999
ACCESSIONS H72266
REFERENCE Nelson, K.E., Clayton, R.A., Gill, S.R., Gwin, M.L., Dodson, J., Nelson, R.J., Haft, D.H., Hickey, E.K., Peterson, J.D., Nelson, W.C., Ketchum, K.A., McDonald, L., Utterback, T.R., Malek, J.A., Linher, K.D., Garrett, M.M., Stewart, A.M., Cotton, M.D., Pratt, M.S., Phillips, C.A., Richardson, D., Heidelberg, J., Sutton, G.G., Fleischmann, R.D., White, O., Salzberg, S.L., Smith, H.O., Venter, J.C., Fraser, C.M., Nature (1999) 399:323-329
Evidence for lateral gene transfer between Archaea and Bacteria from genome sequence of Thermotoga maritima.
Bacteria from genome sequence of Thermotoga maritima.
#journal Nature
#title Evidence for lateral gene transfer between Archaea and Bacteria from genome sequence of Thermotoga maritima.
#accession H72266
#status preliminary
#molecule-type DNA

WISREX (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (C) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MSRCH_PP protein - protein database search, using Smith-Waterman algorithm
Run on: Sat May 13 08:35:40 2000; Maspar time 73.66 Seconds
Tabular output not generated 24.806 Million cell updates/sec

Title: >US-09-331-631-21
Description: (32-91)-from US09331631.pep
Perfect Score: 453

Sequence: 1 TENPCARQRCLOSCQCEPDDL.....DGGATNQRHPGERTGRGP 60

Scoring table: PAM 150
Gap 11

Searched: 83857 seqs, 30454973 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: swiss-prot38
1:swissprot

Statistics: Mean 33.775; Variance 51.458; scale 0.656

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	453	100.0	614	1 AH11_ARAHY	ALLERGEN ARA H 1, CLON	1.78e-92
2	396	87.4	626	1 AH12_ARAHY	ALLERGEN ARA H 1, CLON	1.84e-77
3	120	26.5	639	1 GLCX_SOYBN	BETA-CONGLYCININ, ALPH	1.15e-09
4	119	26.3	605	1 GLCA_SOYBN	BETA-CONGLYCININ, ALPH	1.86e-09
5	93	20.5	680	1 KALM_HUMAN	KALLMANN SYNDROME PROT	2.88e-04
6	85	18.8	676	1 KALM_HUMAN	KALLMANN SYNDROME PROT	8.57e-03
7	77	17.0	4753	1 LRP_GAEEL	LOW-DENSITY LIPOPROTEI	2.14e-01
8	76	16.8	753	1 VCLA_GOSHI	VICILIN GC72-A PRECURS	3.16e-01
9	75	16.6	128	1 RNP_CHOHO	RIBONUCLEASE PANCREATI	4.64e-01
10	75	16.6	211	1 Y121_BURCE	INSERTION ELEMENT IS40	4.64e-01
11	75	16.6	649	1 V70K_EPMV	70 KD PROTEIN,	4.64e-01
12	74	16.3	147	1 IP21_SOLIU	PROTEINASE INHIBITOR T	6.80e-01
13	74	16.3	229	1 VHEL_ECOLI	PROBABLE HELICASE (ORF	6.80e-01
14	74	16.3	320	1 YBIB_ECOLI	HYPOTHETICAL 35.0 KD P	6.80e-01
15	74	16.3	488	1 CRUA_BRANA	CRUCIFERIN PRECURSOR (6.80e-01
16	74	16.3	588	1 VCLB_GOSHI	VICILIN C72 PRECURSOR	6.80e-01
17	74	16.3	1106	1 STC_DROME	SHUTTLE CRAFT PROTEIN.	6.80e-01
18	73	16.1	477	1 ETV1_MOUSE	ERB1 PROTEIN (ETS TRAN	9.92e-01
19	72	15.9	753	1 YJ05_GAEEL	HYPOTHETICAL 85.1 KD N	1.44e+00
20	72	15.9	1955	1 AGRI_CHICK	AGRIN PRECURSOR.	1.44e+00
21	71	15.7	154	1 IP27_SOLIU	PROTEINASE INHIBITOR T	2.09e+00
22	71	15.7	204	1 IP22_CAPAN	WOOND-INDUCED PROTEINA	2.09e+00
23	71	15.7	607	1 YN92_YEAST	POTATIVE TRANSCRIPTION	2.09e+00

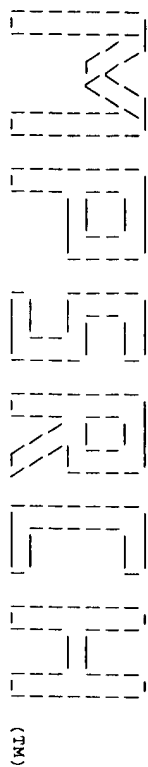
5301D1021

ID	1	STANDARD:	PRT:	614 AA.
AC	P43237			
DT	01-NOV-1995 (Rel. 32, Created)			
DT	01-NOV-1995 (Rel. 32, Last sequence update)			
DT	01-NOV-1997 (Rel. 35, Last annotation update)			
DE	ALLERGEN ARA H 1, CLONE p17 (ARA H I).			
OS	Arachis hypogaea (peanut).			
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;			
OC	euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;			
OC	core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae;			
OC	Arachis.			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN-CV, FLORISSER;			
RX	MEDLINE: 96013631.			
RA	Burks A.W., Cockrell G., Stanley J.S., Helm R.M., Bannan G.A.;			
RT	"Recombinant peanut allergen Ara h I expression and IgE binding in			
RT	patients with peanut hypersensitivity.";			
RL	J Clin Invest. 96:1115-1121(1995).			
CC	- SIMILARITY: TO 75 SEED STORAGE PROTEINS (PHASEOLIN, VICILIN,			
CC	CONVICILIN, CONGLYCININ, ETC.).			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb.ch/announce/			
CC	or send an email to license@isb.slb.ch).			
CC	-----			
CC	EMBL: L38853; AA60336.1; .			
DR	HSSP: P50477; ICAM.			
DR	PFAM: PF00546; Seedstore_7s. 1.			
KW	Allergen.			
SQ	SEQUENCE 614 AA: 70283 MW: 1DDACF217ECC5F31 CRC64;			

Query Match 100.0%; Score 453; DB 1; Length 614;
Best local similarity 100.0%; Pred. No. 1.78e-92;
Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 32 TENPCARQRCLOSCQCEPDDLKACSRCTKLEYDPRCVYDTGATNQRHPGERTGRGP 91
|||||
32 TENPCARQRCLOSCQCEPDDLKACSRCTKLEYDPRCVYDTGATNQRHPGERTGRGP 91

RESULT 2



Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MParch - protein - protein database search, using Smith-Waterman algorithm
Run on: Sat May 13 09:04:49 2000; Maspar time 4.78 seconds
227,008 Million cell updates/sec
Tabular output not generated

Title: >US-09-331-631-25
Description: (1-23) from US09331031.pep
Perfect Score: 177
Sequence: 1 MMRARFPLLLGLVFLASVSYSF 23

Scoring table: PAM 150
Gap 11

Searched: 142080 seqs, 47172406 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: p1r62
1:P1r1 2:P1r2 3:P1r3 4:P1r4

Statistics: Mean 33.374; Variance 80.850; scale 0.413

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	177	100.0	605	1	FMSYBA beta-conglycinin alph	5.92e-13
2	177	100.0	605	2	S20007 beta-conglycinin alph	5.92e-13
3	170	96.0	639	2	B24810 beta-conglycinin alph	6.41e-12
4	138	78.0	421	2	A24810 phaseolin beta chain	2.56e-07
5	138	78.0	430	2	S10156 alpha-phaseolin precu	2.56e-07
6	138	78.0	436	2	A23498 phaseolin type alpha	2.56e-07
7	136	76.8	439	1	FMSYCB beta-conglycinin beta	4.87e-07
8	120	67.8	571	2	S00566 convicillin precursor	7.55e-05
9	114	64.4	296	2	T06572 convicillin precursor	4.75e-04
10	111	62.7	445	2	S00281 canavallin - sword bea	1.18e-03
11	111	62.7	445	2	JO2264 canavallin - jack bean	1.18e-03
12	105	59.3	459	2	S08505 vicillin - garden pea	7.03e-03
13	94	53.1	463	2	A27288 vicillin precursor - f	1.68e-01
14	94	53.1	463	2	S06309 vicillin precursor (cl	2.61e-01
15	84	47.5	547	1	S51475 cytochrome P450 cyp78	2.61e-01
16	83	46.9	438	2	S35757 vicillin, 47k - garden	3.41e-00
17	82	46.3	286	2	G72379 sugar ABC transporter	4.44e-00
18	82	46.3	400	2	F72350 hypothetical protein 2	4.44e-00
19	82	46.3	659	1	B44212 structural protein 2	4.44e-00
20	81	45.8	502	2	G70193 hypothetical protein	5.77e-00
21	81	45.8	909	1	ORXLL1 LDL receptor 1 precur	5.77e-00
22	80	45.2	141	2	E55855 lysa protein - phage	7.49e+00
23	80	45.2	293	2	H72064 metal transport prote	7.49e+00

24	80	45.2	449	2	S16319 secretin receptor - r	7.49e+00
25	79	44.6	67	2	B69936 hypothetical protein	9.70e+00
26	79	44.6	402	2	T13499 NADH dehydrogenase su	9.70e+00
27	79	44.6	405	2	G64971 hypothetical protein	9.70e+00
28	78	44.1	492	2	T02458 hypothetical protein	1.25e+01
29	78	44.1	684	2	T12151 NADH dehydrogenase -	1.25e+01
30	78	44.1	684	2	T13491 NADH dehydrogenase -	1.25e+01
31	78	44.1	685	2	T12129 NADH dehydrogenase -	1.25e+01
32	78	44.1	704	2	T13503 NADH dehydrogenase F	1.25e+01
33	78	44.1	732	2	T14233 NADH dehydrogenase -	1.25e+01
34	78	44.1	734	2	T12602 NADH dehydrogenase -	1.25e+01
35	78	44.1	734	2	T13785 NADH dehydrogenase -	1.25e+01
36	77	43.5	124	2	S45788 probable membrane pro	1.62e+01
37	77	43.5	246	2	E64627 hypothetical protein	1.62e+01
38	77	43.5	271	2	T11692 NADH dehydrogenase su	1.62e+01
39	77	43.5	433	2	G70345 hypothetical protein	1.62e+01
40	77	43.5	660	1	VHMW2 structural protein 2	1.62e+01
41	77	43.5	686	2	T13680 NADH dehydrogenase -	1.62e+01
42	76	42.9	336	2	D71474 probable muramoyl pen	2.09e+01
43	76	42.9	381	2	T13701 NADH dehydrogenase su	2.09e+01
44	76	42.9	534	1	S75101 hypothetical protein	2.09e+01
45	76	42.9	696	2	T13767 NADH dehydrogenase -	2.09e+01

ALIGNMENTS

RESULT 1
ENTRY FMSYBA #type complete
TITLE beta-conglycinin alpha chain precursor - soybean
ORGANISM #formal_name Glycine max #common_name soybean
DATE 31-Dec-1991 #sequence_revision 31-Dec-1991 #text_change 16-Jul-1999

ACCESSIONS S14681; S74124; S06714
REFERENCE S14681
#authors Sebastiani, F.L.; Farrell, L.B.; Schuler, M.A.; Beachy, R.N.
#journal Plant Mol. Biol. (1990) 15:197-201
#title Complete sequence of a cDNA of alpha subunit of soybean beta-conglycinin.

#cross-references MIM:91353680
#accession S14681
#molecule_type mRNA
#residues 1-605 #label SEB

REFERENCE S74123
#cross-references EMBL:X17698; NID:g18535; PIDN:CA35691.1; PID:g18536

#authors Shutoy, A.D.; Kakovskaya, I.A.; Bastrygina, A.S.; Bulmaga, V.P.; Horstmann, C.; Muentz, K.
#journal Eur. J. Biochem. (1996) 241:221-228
#title Limited proteolysis of beta-conglycinin and glycinin, the 7S and 11S storage globulins from soybean [Glycine max (L.) Merr.] - Structural and evolutionary implications.

#cross-references MIM:97054613
#accession S74124
#molecule_type protein
#residues 189-196, 'H', 196, 'N', 200, 'X', 202-203, 397-408, 'X', 410, 'X', 412-417, 'X', 419-420, 'X', 484-501, 'Y', 503 #label SHU

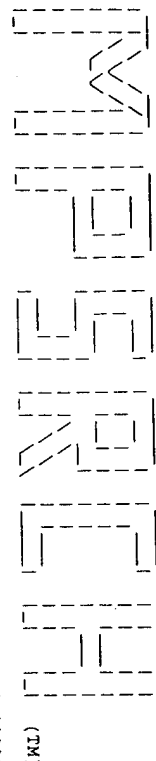
CLASSIFICATION #experimental_source seed
#superfamily glycinin
KEYWORDS glycoprotein; seed; storage protein

FEATURE 1-22
23-62
63-605
261-517
#domain signal sequence #status predicted #label SIG
#domain propeptide #status predicted #label PRO
#product beta-conglycinin alpha chain #status predicted #label MAT
#binding site carbohydrate (Asn) (covalent) #status predicted

SUMMARY #length 605 #molecular_weight 70293 #checksum 6280

Query Match 100.0%; Score 177; DB 1; Length 605;
Best Local Similarity 100.0%; Pred. No. 5.92e-13;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 MMRARFPLLLGLVFLASVSYSF 23
|||||



Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm
Run on: Sat May 13 08:59:23 2000; Maspar time 6.82 Seconds
380.618 Million cell updates/sec

Tabular output not generated.
Title: >US-09-331-631-24
Description: 425
Perfect Score: 425
Sequence: 1 DEDDRRGHSLQCCVORCRQ.....DDQOQHRHDEEGRGRG 55

Scoring table: PAM 150
Gap 11

Searched: 142080 seqs, 47172406 residues
Post-processing: Minimum Match 08
Listing first 45 summaries

Database: p1r62
1:p1r1 2:p1r2 3:p1r3 4:p1r4

Statistics: Mean 32.380; Variance 69.376; scale 0.467

Prod. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description	Pred. No.
1	425	100.0	637	2	S35221	globulin Bgl1 precurs	2.45e-60
2	174	40.9	132	2	C53234	globulin-10 - maize	6.97e-15
3	174	40.9	582	2	B53234	vicilin-like storage	6.97e-15
4	170	40.0	236	2	T01652	globulin-1 - maize (f	3.26e-14
5	168	39.5	407	2	T02258	globulin-1 - maize	7.04e-14
6	168	39.5	540	2	S21825	vicilin-like storage	7.04e-14
7	135	31.8	588	1	FWCNA8	globulin-1S, GUB1S -	1.66e-08
8	135	31.8	588	1	FWCNA8	alpha-globulin B prec	7.06e-08
9	131	30.8	509	2	S08059	alpha-globulin type A	6.23e-04
10	107	25.2	605	2	S06388	hypothetical glycine-	2.37e-03
11	105	24.7	957	2	D70835	trichohyalin - sheep	4.40e-02
12	101	23.8	1549	1	A40691	mannan endo-1,4-beta-	4.40e-02
13	92	21.6	1021	1	T10748	glutactin - fruit fly	4.40e-02
14	92	21.6	1023	1	S12519	trichohyalin - human	1.13e-01
15	92	21.6	1898	1	A45973	vicilin precursor - c	1.13e-01
16	89	20.9	566	2	S22477	hypothetical protein	1.13e-01
17	89	20.9	613	2	S27770	profilaggrin - rat (f	1.13e-01
18	89	20.9	625	2	A34615	hypothetical protein	1.13e-01
19	88	20.7	1200	2	T15921	adenylate cyclase (EC	2.12e-01
20	88	20.7	2248	1	D42088	rlx protein - Staphyl	3.92e-01
21	87	20.5	330	2	S28102	tyrosine kinase activ	3.92e-01
22	85	20.0	450	2	G01158	DNA binding protein -	3.92e-01
23	85	20.0	678	2	S54308		

5321DND:24

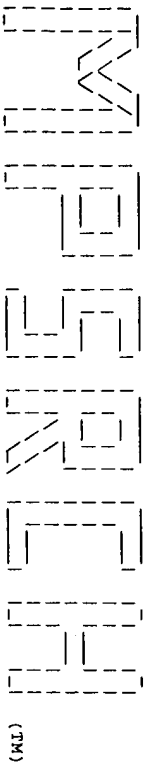
RESULT	ENTRY	ENTRY TITLE	ORGANISM	DATE	ALIGNMENTS
1	S35221	globulin Bgl1 precursor - barley	barley	03-Feb-1994	sequence-revision 03-Feb-1994
2	S35221	globulin Bgl1 precursor - barley	barley	03-Feb-1994	sequence-revision 03-Feb-1994
3	S35221	globulin Bgl1 precursor - barley	barley	03-Feb-1994	sequence-revision 03-Feb-1994
4	S35221	globulin Bgl1 precursor - barley	barley	03-Feb-1994	sequence-revision 03-Feb-1994
5	S35221	globulin Bgl1 precursor - barley	barley	03-Feb-1994	sequence-revision 03-Feb-1994
6	S35221	globulin Bgl1 precursor - barley	barley	03-Feb-1994	sequence-revision 03-Feb-1994
7	S35221	globulin Bgl1 precursor - barley	barley	03-Feb-1994	sequence-revision 03-Feb-1994
8	S35221	globulin Bgl1 precursor - barley	barley	03-Feb-1994	sequence-revision 03-Feb-1994
9	S35221	globulin Bgl1 precursor - barley	barley	03-Feb-1994	sequence-revision 03-Feb-1994
10	S35221	globulin Bgl1 precursor - barley	barley	03-Feb-1994	sequence-revision 03-Feb-1994
11	S35221	globulin Bgl1 precursor - barley	barley	03-Feb-1994	sequence-revision 03-Feb-1994
12	S35221	globulin Bgl1 precursor - barley	barley	03-Feb-1994	sequence-revision 03-Feb-1994
13	S35221	globulin Bgl1 precursor - barley	barley	03-Feb-1994	sequence-revision 03-Feb-1994
14	S35221	globulin Bgl1 precursor - barley	barley	03-Feb-1994	sequence-revision 03-Feb-1994
15	S35221	globulin Bgl1 precursor - barley	barley	03-Feb-1994	sequence-revision 03-Feb-1994
16	S35221	globulin Bgl1 precursor - barley	barley	03-Feb-1994	sequence-revision 03-Feb-1994
17	S35221	globulin Bgl1 precursor - barley	barley	03-Feb-1994	sequence-revision 03-Feb-1994
18	S35221	globulin Bgl1 precursor - barley	barley	03-Feb-1994	sequence-revision 03-Feb-1994
19	S35221	globulin Bgl1 precursor - barley	barley	03-Feb-1994	sequence-revision 03-Feb-1994
20	S35221	globulin Bgl1 precursor - barley	barley	03-Feb-1994	sequence-revision 03-Feb-1994
21	S35221	globulin Bgl1 precursor - barley	barley	03-Feb-1994	sequence-revision 03-Feb-1994
22	S35221	globulin Bgl1 precursor - barley	barley	03-Feb-1994	sequence-revision 03-Feb-1994
23	S35221	globulin Bgl1 precursor - barley	barley	03-Feb-1994	sequence-revision 03-Feb-1994

ACCESSIONS
S35221
REFERENCE
Heck, G.R.; Chamberlain, A.K.; Ho, T.H.D.
Mol. Gen. Genet. (1993) 259:209-218
Barley embryo globulin I gene, Bgl1: characterization of
cDNA, chromosome mapping and regulation of expression.
#cross-references MIM:93287988
#accession S35221
#molecule-type mRNA
#residues 1-637
#label HEC
#cross-references EMBL:M64372; NID:g167003; PID:g167004

GENETICS
#gene Bgl1
#map position 4
#classification superfamily glycinin
#keywords glycoprotein
FEATURE
174-190
SUMMARY
#product globulin Bgl1
#status predicted
#label MAT
#length 637
#molecular-weight 72252
#checksum 3309

Query Match
Best Local Similarity 100.0%; Score 425; DB 2; Length 637;
Matches 55; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
DB 31 DEDDRRGHSLQCCVORCRQERPRYSARCVORCDQDOOHGRHDEEGRGRG 85

RESULT 2
ENTRY C53234
TITLE globulin-10 - maize
ORGANISM Zea mays
DATE 02-May-1994
ACCESSIONS C53234; S21823
REFERENCES Belanger, F.C.; Kriz, A.L.
Genetics (1991) 129:863-872
Molecular basis for allelic polymorphism of the maize
globulin-1 gene.



Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

Wpserch_pp protein - protein database search, using Smith-Waterman algorithm
Run on: Sat May 13 08:40:44 2000: Maspar time 96.60 Seconds
18.915 Million cell updates/sec

Tabular output not generated

Title: >US-09-331-631-22
Description: (45-RA) from HSD931631.pep
Perfect Score: 459
Sequence: 1 EDNHHHGGHKSQCVRRC.....EKROERSRHEADRSRGSS 60

Scoring table: PAM 150
Gap 11

Searched: 83857 seqs, 30454973 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: swiss-prot38
1:swissprot

Statistics: Mean 34.513: Variance 68.998: scale 0.500

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description	Pred. No.
1	453	98.7	573	1	GLB1_MAIZE	GLOBULIN-1 S ALLELE PR	1.62e-09
2	106	23.1	605	1	VCLB_GOSHI	VICILIN GCT2-A PRECURS	2.85e-04
3	103	22.4	1407	1	TRHY_RABIT	TRICHOHYALIN	8.10e-04
4	96	20.9	380	1	RDP_HUMAN	RD PROTEIN	8.82e-03
5	96	20.9	588	1	VCLB_GOSHI	VICILIN GCT2 PRECURSOR	8.82e-03
6	95	20.7	412	1	YNP1_CAEEL	HYPOHETICAL 49.5 KD P	1.23e-02
7	95	20.7	1200	1	DXB8_CAEEL	PUTATIVE PRE-MRNA SPLI	1.23e-02
8	95	20.7	1998	1	TRHY_HUMAN	TRICHOHYALIN	1.23e-02
9	94	20.5	632	1	CCB2_RABIT	DIHYDROXYRIDINE-SENSIT	1.72e-02
10	92	20.0	241	1	YOB7_CAEEL	HYPOTHETICAL 28.5 KD P	3.33e-02
11	92	20.0	466	1	CYP8_CAEEL	PEPTIDYL-PROLYL CIS-TR	3.33e-02
12	91	19.8	373	1	MBP1_MAIZE	ANTIMICROBIAL PEPTIDE	4.63e-02
13	91	19.8	375	1	RDP_MOUSE	RD PROTEIN (W623)	6.41e-02
14	90	19.6	470	1	LEU2_ACOVI	LEGUMIN A PRECURSOR	6.41e-02
15	90	19.6	517	1	LEGA_PEA	LEGUMIN A2 PRECURSOR	6.41e-02
16	90	19.6	520	1	LEGA_PEA	LEGUMIN A2 PRECURSOR	6.41e-02
17	90	19.6	722	1	HMN2_DROME	HOMEOBOX PROTEIN VND (6.41e-02
18	90	19.6	2327	1	CCAB_MOUSE	VOLTAJE-DEPENDENT N-TY	6.41e-02
19	89	19.4	407	1	IE68_HSVSA	IMMEDIATE-EARLY PROTEI	8.87e-02
20	89	19.4	471	1	RUI7_XENLA	U1 SMALL NUCLEAR RIBON	8.87e-02
21	87	19.0	1157	1	Y182_HUMAN	HYPOTHETICAL PROTEIN K	1.69e-01
22	87	19.0	1335	1	XDH_DROME	XANTHINE DEHYDROGENASE	1.69e-01
23	86	18.7	2336	1	CCAB_RAT	VOLTAJE-DEPENDENT N-TY	2.32e-01

SEQ ID NO: 22

24	85	18.5	196	1	SFR2_CAEEL	PUTATIVE SPLICING FACT	3.18e-01
25	85	18.5	197	1	TRF2_DROME	FEMALE-SPECIFIC TRANSF	3.18e-01
26	85	18.5	344	1	SFR6_HUMAN	SPLICING FACTOR, ARGIN	3.18e-01
27	85	18.5	668	1	F13B_MOUSE	COAGULATION FACTOR XII	3.18e-01
28	85	18.5	815	1	MK07_HUMAN	MITOGEN-ACTIVATED PROT	3.18e-01
29	85	18.5	935	1	RNE_HAEIN	RIBONUCLEASE E (EC 3.1	3.18e-01
30	84	18.3	382	1	NCAP_CVCAE	NUCLEOCAPSID PROTEIN	4.38e-01
31	84	18.3	1121	1	DDX8_ARATH	PUTATIVE PRE-MRNA SPLI	4.38e-01
32	83	18.1	285	1	YD41_MYCE	HYPOTHETICAL 30.9 KD P	5.95e-01
33	83	18.1	494	1	SFR4_HUMAN	SPLICING FACTOR, ARGIN	5.95e-01
34	82	17.9	416	1	U2AF_DROME	SPLICING FACTOR U2AF 5	8.12e-01
35	82	17.9	503	1	YOM5_CAEEL	HYPOTHETICAL 57.1 KD P	8.12e-01
36	82	17.9	661	1	F13B_HUMAN	COAGULATION FACTOR XII	8.12e-01
37	82	17.9	1106	1	STC_DROME	SHUTTLE CRAFT PROTEIN	8.12e-01
38	82	17.9	1227	1	PR16_HUMAN	PRE-MRNA SPLICING FACT	8.12e-01
39	81	17.6	437	1	YG3_HALSO	HYPOTHETICAL 50.6 KD P	1.10e+00
40	81	17.6	516	1	LEGB_GOSHI	LEGUMIN B PRECURSOR (B	1.10e+00
41	81	17.6	524	1	SBP_SOYBN	SUCROSE-BINDING PROTEI	1.10e+00
42	81	17.6	1247	1	POL5_ONNNG	STRUCTURAL POLYPROTEIN	1.10e+00
43	80	17.4	587	1	ESR1_POEGU	ESTROGEN RECEPTOR (ER)	1.50e+00
44	80	17.4	612	1	DY12_MOUSE	DYNEIN INTERMEDIATE CH	1.50e+00
45	80	17.4	1157	1	SRA4_HUMAN	CTD-BINDING SR-LIKE PR	1.50e+00

ALIGNMENTS

RESULT ID	1	STANDARD:	PRT:	573 AA.
AC	P15590:			
DT	01-APR-1990 (Rel. 14, Created)			
DT	01-AUG-1990 (Rel. 15, Last sequence update)			
DT	15-JUL-1999 (Rel. 38, Last annotation update)			
DE	GLOBULIN-1 S ALLELE PRECURSOR (GLB1-S) (7S-LIKE).			
GN	GLB1.			
OS	Zea mays (Maize).			
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;			
OC	euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales;			
OC	Poaceae; Zea.			

(1)
SEQUENCE FROM N.A.
RC STRAIN-CV. INBRED LINE VA26;
RA Belanger F.C., Kriz A.L.;
KT "Molecular characterization of the major maize embryo globulin encoded by the GLB1 gene."
RL Plant Physiol. 91:636-643(1989).
(2)
SEQUENCE OF 87-100.
RP MEDLINE: 89374022.
RX
RA Kriz A.L.;
RT "Characterization of embryo globulins encoded by the maize Glb1 genes."
RT Biochem. Genet. 27:239-251(1989).
RT "PM: THREE PROTEIN-PROCESSING STEPS OCCUR IN THE FORMATION OF THE MATURE PROTEIN FROM THE PRIMARY TRANSLATION PRODUCT."
CC - POLYMORPHISM: THE THREE MOST COMMONLY OCCURRING GLB1 ALLELES HAVE THE DESIGNATION L, I, AND S FOR LARGE, INTERMEDIATE, AND SMALL PROTEINS, RESPECTIVELY.
CC - SIMILARITY: TO OTHER 7S SEED STORAGE PROTEINS (PHASEOLIN, VICILIN, CONVICILIN, CONGLUTININ, ETC.).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use - by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC EMBL: M248453 AAA33467.1;
CC HSSP: P50477; ICAM.
CC MAZEDB: 30181;
CC PIRAM: PF00546; Seedstore_7s; 1.
CC Seed storage protein; signal.

WPISEED

(TM)

Release 3.1A John F. Collins. Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MSrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Sat May 13 08:42:53 2000; Maspar time 167.03 Seconds
24.906 Million cell updates/sec

Tabular output not generated.

Title: 5US-09-331-631-22
(25-047-444-450-331-631-22)

Description: 459
Perfect Score: 1 EDDNHHHGHKSGCVCRRRC.....EKROENRHEADRSGEGSS 60
Sequence:

Scoring table: PAM 150
Gap 11

Searched: 225878 seqs, 69334122 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: sptrembl12
1:sp:archaea 2:sp:bacteria 3:sp:fungi 4:sp:human
5:sp:invertebrate 6:sp:mammal 7:sp:mhc 8:sp:organelle
9:sp:phage 10:sp:plant 11:sp:rodent 12:sp:unclassified
13:sp:vertebrate 14:sp:virus

Statistics: Mean 33.511; Variance 71.985; scale 0.466

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description	Pred. No.
1	459	100.0	122	10	GLOBULIN-1 0 ALLELE PR	9.08e-66
2	459	100.0	582	10	GLOBULIN-1 L ALLELE PR	9.08e-66
3	459	98.7	240	10	GLOBULIN-1 (FRAGMENT)	1.22e-64
4	453	98.7	540	10	GLOBULIN-1 S ALLELE PR	1.22e-64
5	450	98.0	238	10	GLOBULIN-1 (FRAGMENT)	4.50e-64
6	446	97.2	238	10	GLOBULIN-1 (FRAGMENT)	2.54e-63
7	446	97.2	238	10	GLOBULIN-1 (FRAGMENT)	2.54e-63
8	446	97.2	407	10	GLOBULIN-1 (FRAGMENT)	3.92e-63
9	445	96.9	238	10	GLOBULIN-1 (FRAGMENT)	9.11e-63
10	443	96.5	240	10	GLOBULIN-1 (FRAGMENT)	2.21e-62
11	441	96.1	242	10	GLOBULIN-1 (FRAGMENT)	3.42e-59
12	424	92.4	236	10	GLOBULIN-1 (FRAGMENT)	4.64e-50
13	424	92.4	239	10	GLOBULIN-1 (FRAGMENT)	4.64e-50
14	375	81.7	246	10	GLOBULIN-1 (FRAGMENT)	2.15e-48
15	366	79.7	242	10	GLOBULIN-1 (FRAGMENT)	1.05e-44
16	346	75.4	236	10	GLOBULIN-1 (FRAGMENT)	3.75e-43
17	343	74.7	242	10	GLOBULIN-1 (FRAGMENT)	3.03e-15
18	180	39.2	637	10	EMBRIO GLOBULIN	6.82e-05
19	115	25.1	393	10	HYPOTHETICAL 45.3 KD P	1.01e-03
20	107	23.3	402	10	EARLY EMBRYOGENESIS PR	

SCAID NO.22

21	103	22.4	238	5	THROMBOSPONDIN-RELATED	3.79e-03
22	99	21.6	204	13	SBCFR-1 PROTEIN PRECUR	1.39e-02
23	97	21.1	444	5	F56A11.6 PROTEIN.	2.65e-02
24	97	21.1	672	5	F28C1.1 PROTEIN.	2.65e-02
25	96	20.9	453	2	PLASMIN PRLI. COMPLETE	3.64e-02
26	96	20.5	402	4	CGI-74 PROTEIN.	6.86e-02
27	92	20.0	555	10	U2 SNRP AUXILIARY FAC	1.28e-01
28	91	19.8	292	10	SMALL NUCLEAR RIBONUCL	1.75e-01
29	91	19.8	342	11	MHC CLASS III REGION R	1.75e-01
30	91	19.8	450	4	TYROSINE KINASE ACTIVA	1.75e-01
31	91	19.8	499	10	LEGUMIN (FRAGMENT).	1.75e-01
32	90	19.6	155	5	ROSD11.7 PROTEIN.	2.39e-01
33	90	19.6	897	5	LET 858.	2.39e-01
34	90	19.6	1104	13	SERINE-THREONINE KINAS	2.39e-01
35	89	19.4	163	4	NUCLEIC ACID BINDING P	3.25e-01
36	89	19.4	337	2	PURATIVE COINTEGRATE R	3.25e-01
37	89	19.4	405	4	45KDA SPLICING FACTOR.	3.25e-01
38	89	19.4	630	5	COSMID K03H11.	3.25e-01
39	89	19.4	927	2	INITIATION FACTOR IF2.	3.25e-01
40	89	19.4	997	5	CODED FOR BY C. ELEGAN	3.25e-01
41	89	19.4	1276	10	SIMILAR TO AKONEME-ASS	3.25e-01
42	88	19.2	179	5	SIMILAR TO PRE-MRNA SP	4.41e-01
43	88	19.2	515	5	W06F12.2 PROTEIN	4.41e-01
44	88	19.2	573	10	U2 SNRP AUXILIARY FAC	4.41e-01
45	87	19.0	606	4	CALCIUM CHANNEL BETA2A	5.98e-01

ALIGNMENTS

RESULT	ID	1	PRELIMINARY:	PRT:	122 AA.
AC	003863				
DT	01-NOV-1996	(TREMBLrel. 01, Created)			
DT	01-NOV-1996	(TREMBLrel. 01, Last sequence update)			
DT	01-NOV-1999	(TREMBLrel. 12, Last annotation update)			
DE	GLOBULIN-1 0 ALLELE PRECURSOR (GLB1-0) (7S-LIKE).				
GN	GLB1-0.				
OS	Zea mays (Maize).				
OC	Eukaryota: Viridiplantae: Streptophyta: Embryophyta: Tracheophyta:				
OC	euphyllophytes: Spermatophyta: Magnoliophyta: Liliopsida: Poales:				
OC	Poaceae; Zea.				
RN	(1)				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE: 92090707.				
RA	BELANGER F.C., KRIZ A.L.;				
RT	*Molecular basis for allelic polymorphism of the maize globulin-1				
RL	gene.				
CC	Genetics 129:863-872(1991)				
CC	MISCELLANEOUS: THREE PROTEIN-PROCESSING STEPS OCCUR IN THE				
CC	FORMATION OF THE MATURE PROTEIN FROM THE PRIMARY TRANSLATION				
CC	PRODUCT.				
CC	-1- SIMILARITY: TO OTHER 7S SEED STORAGE PROTEINS (PHASEOLIN, VICILIN,				
CC	CONVICTILIN, CONGLYCININ, ETC.).				
DR	EMBL: X59085; CAA41811.1;				
DR	MEDEL: 14637; Zema:1188;14637.				
KW	Seed storage protein.				
FT	SIGNAL				
FT	CHAIN				
FT	SEQUENCE				
SO	122 AA: 13902 MW: 89C81691 CRC32:				
Query Match	100.0%:	Score 459:	DB 10:	Length 122:	
Best Local Similarity 100.0%:	Pred. No. 9.08e-66:				
Matches 60:	Conservative 0:	Mismatches 0:	Indels 0:	Gaps 0:	
DB	25 EDDNHHHGHKSGCVCRRCDPRHMHQPRCLDCCREERKROERSHEDRSGEGSS 84				
Oy	25 EDDNHHHGHKSGCVCRRCDPRHMHQPRCLDCCREERKROERSHEDRSGEGSS 84				
RESULT	2				
ID	003865				
AC	003865				
DT	01-NOV-1996	(TREMBLrel. 01, Created)			

WISDOM (TM)

Release 3.1A John F. Collins, BioComputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

March-pp protein - protein database search, using Smith-Waterman algorithm
Run on: Sat May 13 08:46:12 2000; MasPar time 6.97 Seconds
406.080 Million cell updates/sec

Tabular output not generated

Title: >US-09-331-631-22

Description: (25-84) from US09331631.pep

Perfect Score: 459

Sequence: 1 EDDNNHHGHGKSGQCVRC.....EKQERSHEDDRSGEGSS 60

Scoring table: PAM 150
Gap 11

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database: p1r62
1:p1r1 2:p1r2 3:p1r3 4:p1r4

Statistics: Mean 33.532; Variance 75.335; scale 0.445

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description	Pred. No.
1	459	100.0	122	2	C53234	globulin-10 - maize	5.93e-63
2	459	100.0	582	2	B53234	vicillin-like storage	5.93e-63
3	453	98.7	540	2	S21825	vicillin-like storage	7.14e-62
4	453	98.7	573	2	A53234	globulin-1S, GLB1S -	7.14e-62
5	446	97.2	407	2	T02258	globulin-1 - maize	1.30e-60
6	343	74.7	236	2	T01662	globulin-1 - maize (f	2.79e-42
7	180	39.2	637	2	S35221	globulin Bgl precurs	1.25e-14
8	106	23.1	605	2	S06398	alpha-globulin type A	1.83e-03
9	103	22.4	1407	1	S28589	trichohyalin - rabbit	4.72e-03
10	96	20.9	325	2	JH0189	arginine/aspartate-ti	4.12e-02
11	96	20.9	380	2	S36789	gene RD protein - hum	4.12e-02
12	96	20.9	382	2	A33640	class III histocompat	4.12e-02
13	96	20.9	588	1	FMCNAB	alpha-globulin B prec	4.12e-02
14	95	20.7	346	2	S41001	hypothetical protein	5.58e-02
15	95	20.7	1200	2	T15921	hypothetical protein	5.58e-02
16	95	20.7	1898	1	A45973	trichohyalin - human	5.58e-02
17	94	20.5	606	2	S21046	calcium channel prote	7.55e-02
18	94	20.5	632	2	S21048	calcium channel prote	7.55e-02
19	92	20.0	241	2	S44893	zx1236.7 protein - Ca	1.38e-01
20	92	20.0	466	2	T15877	hypothetical protein	1.38e-01
21	91	19.8	33	2	A41822	antimicrobial peptide	1.85e-01
22	91	19.8	375	2	A40112	MHC-region RD-repeat	1.85e-01
23	91	19.8	450	2	G01158	lysine kinase activ	1.85e-01

24	90	19.6	203	2	S09559	legumin A (clone pCD4	2.49e-01
25	90	19.6	507	2	T06452	probable legumin A pr	2.49e-01
26	90	19.6	517	1	FPFMLA	legumin A precursor -	2.49e-01
27	90	19.6	520	2	S08237	legumin A2 precursor	2.49e-01
28	90	19.6	722	2	S57246	ventral nervous syste	2.49e-01
29	89	19.4	163	2	I38191	nucleic acid binding	3.35e-01
30	89	19.4	407	1	EDBEQ3	immediate-early prote	3.35e-01
31	89	19.4	471	2	S02016	U1 snRNP 70K protein	3.35e-01
32	89	19.4	1276	2	T02711	hypothetical protein	3.35e-01
33	87	19.0	1335	2	S07245	xanthine dehydrogenas	6.01e-01
34	86	18.7	509	2	S08059	alpha-globulin type B	8.03e-01
35	86	18.7	1380	2	T00365	hypothetical protein	8.03e-01
36	86	18.7	1473	2	T13855	suppressor of sable p	8.03e-01
37	86	18.7	2336	2	A45386	omega-conotoxin-sens1	8.03e-01
38	85	18.5	196	2	T15917	hypothetical protein	1.07e+00
39	85	18.5	197	2	A29648	female-specific trans	1.07e+00
40	85	18.5	344	2	S59043	spilling factor SRP55	1.07e+00
41	85	18.5	566	2	S22477	vicillin precursor - c	1.07e+00
42	85	18.5	668	2	A46013	coagulation factor XI	1.07e+00
43	85	18.5	815	2	B56708	extracellular signal-	1.07e+00
44	85	18.5	951	1	E64066	ribonuclease E (EC 3.	1.07e+00
45	85	18.5	2339	2	A42566	omega-conotoxin-sens1	1.07e+00

ALIGNMENTS

RESULT 1	C53234	*type complete
ENTRY	globulin-10 - maize	
TITLE	*formal_name 2ea may5 *common_name maize	
ORGANISM	02-May-1994 *sequence_revision 18-Nov-1994 *text_change	
DATE	10-Oct-1997	
ACCESSIONS	C53234; S21823	
REFERENCE	A53234	
*authors	Belanger, F.C.; Kriz, A.L.	
*journal	Genetics (1991) 129:863-872	
*title	Molecular basis for allelic polymorphism of the maize	
	globulin-1 gene.	
*cross-references	MUID:92090707	
*contents	Black Beauty popcorn	
*accession	C53234	
*status	preliminary	
**molecule_type	DNA	
**residues	1-122 *label BEL	
**cross-references	EMBL:X59085; NID:g22281; PID:g22282	
*note	Sequence extracted from NCBI backbone (NCBIN:71282, NCBI:P:71286)	

SUMMARY #length 122 #molecular_weight 13902 #checksum 4000

Query Match 100.0%; Score 459; DB 2; Length 122;

Best Local Similarity 100.0%; Pred. No. 5.93e-63;

Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB	25	EDDNNHHGHGKSGQCVRCEDRPHORPCLCEOCREERKROERSRHEADDSGEGSS 84
Qy	25	EDDNNHHGHGKSGQCVRCEDRPHORPCLCEOCREERKROERSRHEADDSGEGSS 84
ENTRY	2	B53234 *type complete
TITLE		vicillin-like storage protein Gbl-1, embryo - maize
ALTERNATE_NAMES		*formal_name 2ea may5 *common_name maize
ORGANISM		02-May-1994 *sequence_revision 18-Nov-1994 *text_change
DATE		09-Sep-1997
ACCESSIONS		B53234; S21824
REFERENCE		A53234
*authors		Belanger, F.C.; Kriz, A.L.
*journal		Genetics (1991) 129:863-872
*title		Molecular basis for allelic polymorphism of the maize
		globulin-1 gene.
*cross-references		MUID:92090707
*accession		B53234